University of Alberta

A Posthuman Investigation: Assessing the Suitability of Consciousness to Digital Duplication

by

Olaf Ellefson

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the

requirements for the degree of *Master of Arts*

in

Philosophy

Humanities Computing

Edmonton, Alberta
Spring 2006

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-494-13730-4
Our file Notre référence
ISBN: 0-494-13730-4

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

This thesis aims to assess the claims of posthumanist scholars such as Moravec and Tipler who assert that human consciousness will one day be transferable into digital computers. The first goal of this project is to explore consciousness via a Searlean perspective (i.e., a perspective that includes subjectivity and qualia) as it poses serious challenges to a digital account of consciousness. The second goal is to explore computation and the relationship it might have to consciousness and intentionality. This involves, among other things, responding to Searle's characterization of computation as observer-relative with McDermott's account of an objective computer science. Ultimately, I argue for consciousness to remain as Searle views is – a phenomenon rooted in the biology and body of human beings – but also for an acknowledgement that computation plays a significant role in that biology.

Table of Contents

0.0	A Posthuman Investigation: Assessing the Suitability of Consciousness to Digitial Duplication	1
1.0	Consciousness	4
1.1	Introduction	4
1.2	Qualities of Consciousness	8
1.2.1	Subjectivity	9
1.2.2	Qualia	14
1.2.3	Unity	22
1.3	Conclusion	26
2.0	Computation	28
2.1	Against Strong AI: The Chinese Room Argument	28
2.1.1	Intentionality	30
2.1.2	The Intentional Stance	32
2.1.3	The Systems Reply	36
2.2	Against Computationalism: Defining Computation	43
2.2.1	McDermott's Reply	45
2.2.2	Diagnosis and Conclusion	50
3.0	Conclusion	53
4.0	References	56

List of Figures

Figure 2.1. Computational causality 47

0.0 A Posthuman Investigation: Assessing the Suitability of Consciousness to Digital Duplication

Posthumanism, briefly, is a theory that expounds on the possibility of human beings (or their descendants) creating sufficiently powerful computers to enable the full transference or duplication of their carnal existence into digital media. The ideas behind posthumanism have been common for years throughout many sub-genres of science fiction,[1] but only more recently have they begun to make headway in mainstream and academic thought. Such progress can, I think, be attributed in part to the writings of Turkle and Hayles and their respective explorations of digital lifestyles and cultural metaphors surrounding new media. However, nowhere is a posthuman agenda more evident than in Tipler's *The Physics of Immortality* (1994) in which he envisions a future where all past and present human beings are resurrected into a universal virtual reality device to live in perpetual peace and utopia. But others such as Bostrom (2003) have explored the probability of civilizations with ancestor simulation capabilities actually coming to be and he startlingly concludes that if we are not already in a simulation, then it is extremely unlikely that we ever will be.

Clearly, there is a great deal of discussion and debate (no matter how fanciful) surrounding posthumanism and much time has been spent imagining the intricacies of such an existence. Yet what has rarely been addressed is that which is most central to the likelihood of posthumanism's success: what is the role of biology in human consciousness and can that role be managed by computational systems? After all, a posthuman project that would only create a zombie or a human "vegetable", regardless of their impressive construction or behaviour, is ultimately a failure. While I cannot, and am unwilling, to answer the question "What is human?" I think I can explore elements of humanity that would be central to posthumanism's success, such as consciousness and intentionality. Defining these properties is difficult as they are particularly elusive, but an attempt must be made as they are nonetheless integral to the "human experience", and equally so for a posthuman project to be considered successful or desirable.

---

1 See, for instance, Leonard's film *The Lawnmower Man* (1994), or Greg Egan's novel *Diaspora* (1999).

It is important at this point that I draw a distinction between transhumanism, the theory that humanity should foster the development of technologies that will alter our biology and augment our physical and mental abilities through prosthetics and implants, and the absolute abandonment of the body for residence within a virtual world, which is posthumanism. Both theories are intimately related and promise revolutionary and exciting possibilities for humanity and, in doing so, raise very difficult questions about what it is to be human and our place in the world. My focus will be on posthumanism as transhuman projects, to varying degrees of completion, are already well underway (e.g., vaccinations, prosthetic limbs, microchip implants, and so on). While these programs are undeniably interesting, even if some of their merits are debatable, they remain unsuitable for my project as they do not, in the same manner as posthumanism, ask us to investigate what we are. For with transhumanism, the body, and our embodiment within its biological frame, is taken as a given and it is only our abilities that are altered and enhanced, but with posthumanism the body as an organic construct is entirely stripped away and (human) consciousness is to be recreated in an entirely new, non-biological form.

Accordingly, if posthumanism is to be successful in duplicating the full range of human experience, it must address beforehand what it is that enables those very experiences, and ask, as does Dennett (1991a), what and how important they are. Such questions will be addressed in the first section of this project and, due to the immense task of determining the role of biology in consciousness and intentionality, I have chosen to focus on a few of the qualities of consciousness, as defined by Searle (2000), which seem to be particularly tied to our physiology. Of course, the study of consciousness is highly contested and the methodologies to do so are often under constant redefinition and reinterpretation. Accordingly, I aim not to cement a definition of consciousness, but to explore and examine a few of the properties that allow for human experience, with an eye to those that must be present for a posthuman project to be realized.

The second part of my project will ask, once it is known what needs to be duplicated, if such a thing can be performed by a computer. This question naturally raises further questions of what designates and defines a computer, in addition to those regarding the properties of human minds. Significant and compelling work has been done

in this area and so I will, in the second section, begin my investigation with Searle's Chinese Room (1980, 1984), which, in addition to his further critiques of computationalism, is amongst the strongest arguments against artificial intelligence. Searle hopes to show that computer simulations of intelligence are missing key human abilities, namely intentionality and consciousness, and so, for my purposes, are insufficient to guarantee a means to a credible posthuman project. Searle is not, however, the final word on this topic and, as his argument against computationalism pertains to a very specific definition of a computer, it is debatable if his argument can withstand newer conceptions of computation and computers. As such, the probability of success for posthumanism is still unknown and it is this question, above all others, that I hope to explore and to which I might provide an answer.

1.0 Consciousness

1.1 Introduction

Searle, a well-known critic of artificial intelligence, would seem to be an unlikely choice for someone beginning an investigation into the posthuman. However, as Searle presents rigourous examinations of consciousness in the context of artificial intelligence and computer science research, he is nonetheless a relevant, if unorthodox, choice. This is not to say that my treatment of Searle should be taken as the final word in regards to his philosophy. In my proceeding attempt to evaluate the possibility for an authentic posthumanism, I will be pushing Searle's work in a direction he would undoubtedly be disinclined to support himself. So, too, with Dennett; as an unequivocal supporter of computer intelligence, it would seem as if I should be fully endorsing his views and not Searle's. Yet this is not, as will be seen, what I end up doing. Instead, by investigating the antagonisms between Searle and Dennett, I hope to show that posthumanism is possible, although not, perhaps, possible in a way that either philosopher would expect (or even like). My goal is not to provide a traditional defence of either philosopher's work, but is to avoid often uncritically optimistic visions of posthumanism. Such visions, as seen in works by Tipler and Moravec, see the move from human to post as being, if not trivial, then at least unproblematic in that equivalence between computers and human brains is assumed to be (eventually) guaranteed. Such an assumption is usually made in the absence of a deeper study of consciousness. To address this absence, I begin my project with an exploration of consciousness from a Searlean perspective with a Dennettian twist. Each philosopher has fundamentally different conceptions of consciousness: Searle affirms its existence, Dennett denies/redefines it. Such a discrepancy cannot go unnoticed and, as it is my belief that such incongruities often lead to fruitful discussions, a comparison of their works should prove helpful.

Searle begins his investigation into consciousness against a historical backdrop of the philosophy of mind; in so doing he formulates a solution (or dissolution as he might call it) to the mind/body problem, which takes the form of what he calls "biological naturalism". In its shortest form, it is simply the hypothesis that "[m]ental phenomena are caused by neurophysical processes in the brain and are themselves features of the

brain," (Searle 1992, p.1). At first glance, this hypothesis does not appear to be a radical break from more traditional theories of mind such as reductivism and materialism. However, as will be shown, Searle views consciousness as causally dependent upon the brain, but physically irreducible to it. That the mind is dependent upon the brain is unsurprising and uncontroversial in most circles (although dualists and proponents of strong AI might disagree), but what is odd is Searle's claim that consciousness cannot be reduced to physics. To illustrate, Searle's comparison of digestion and cognition is helpful. Just as the stomach has the potential to bring about digestion, so too does the brain bring about cognition – both are emergent biological phenomena and nothing mystical or ineffable is involved in their production. Nonetheless, an important distinction must be made between the processes of cognition and digestion as the latter can be described entirely in third-person accounts – there is nothing subjective involved in digestion – while such a description is unavailable to the former as any such analysis would not capture the subjectivity of consciousness. A dissertation on neuronal activity does not adequately express what it is *like* to be happy, sad, or to be thinking about the weather. Searle, then, is neither a dualist nor a materialist as he affirms the existence of conscious, subjective experiences that are dependent upon physics but, because of subjectivity, denies that they are entirely reducible to it.

Additionally, he sharply criticizes research projects that have tried to reduce/ naturalize mental phenomena such as intentionality and consciousness to brute physical processes. Such projects, he states, are doomed because they leave out the fact that intentional and conscious states are always *someone's* intentional and conscious states (ibid., p. 20). Materialist projects have consistently made this mistake because they have confused notions of ontological objectivity and subjectivity with epistemic objectivity and subjectivity. Materialists, of which scientists are a sort, are concerned with phenomena that can be known objectively and are trained to ignore subjective information. For instance, some scientists might investigate the number of species of birds in the world and it is, of course, an objectively ontological fact that *n* bird species exist and scientists can access this knowledge by carrying out scientific research, which follows methodologies that can be agreed upon by a wide range of other scholars. It would be irrelevant to the study if the scientists were to record the number of species

based on their own personal, subjective interests (e.g., by the birds they liked and did not like) and scientists are, of course, trained not to do precisely that.

The problem, Searle states, is that in the desire of scientists and materialists to maximize their objectivity, they began denying or trivializing the existence of all phenomena that hint of any variety of subjectivity (ibid., p. 19-20). Consciousness, feelings, experiences, and intentionality are all subjective phenomena – they exist for one person and are not verifiable in the same way that the number of bird species is – and so scientists have denied their importance in order to maintain the doctrine of (epistemic) objectivity. Consequently, scientists have incorrectly assumed that the only objects worthy of proper investigation are ontologically objective phenomena because they do not wish to be seen studying non-scientific phenomena, which stems from their inability to distinguish between the ontological and the epistemic. To clarify, there are a variety of phenomena in the world that are either ontologically objective (e.g., gravity, mass, number) or subjective (e.g., consciousness, intentionality). These phenomena exist regardless of how we come to know them, whether it is through objective, third-person methodologies, or more subjective accounts, such as those characterized by personal bias or prejudice. However, once objective epistemic methodologies were privileged over the subjective, a similar, though unnecessary, split was made in ontological classifications of the world. The end result is that those who are interested in consciousness have, until recently, been cast in an unscientific or dualist light and consciousness itself is either denied existence or, when it is discussed, is reinterpreted in such a way as to be utterly unfamiliar.[2] But this need not be the case as the irreducibility of consciousness to third-person accounts is resultant from our theoretical methodologies and not from any deep problem about consciousness. Of this, Searle states:

> Consciousness fails to be reducible not because of some mysterious feature, but simply because by definition it falls outside the pattern of reduction we have chosen to use for pragmatic reasons. Pretheoretically, consciousness, like solidity, is a surface feature of certain physical systems. But unlike solidity, consciousness cannot be redefined in terms of an underlying microstructure, and the surface features then treated as mere effects of real consciousness, without losing the point of having the concept of consciousness in the first place. (ibid.,p. 122-23)

---

2 This theme will reemerge in the following discussion of Dennett.

In other words, we can redefine heat as molecular movement and treat the feeling of warmth as a byproduct of that movement, just as we can redefine a sunset as an optical illusion and rest assured that the sun does not orbit the Earth. However, when it comes to consciousness, what it is to be conscious is to have subjective experiences; to focus on reducing those experiences to third-person objective phenomena denies the centrality of those qualities to consciousness. Such analyses give accounts of the forces involved in the production of consciousness and their usefulness cannot be underestimated, but they do not explain consciousness-as-a-subjective-experience and so do not explain consciousness itself. Unsurprisingly, this oversight has lead to the proliferation of behaviouristic approaches in psychology and cognitive science, of which the Turing Test is arguably the most famous. The problem with these approaches is that they get the question wrong; it is not "How can I tell if *x* is intelligent?", but "How can I tell if *x* is thinking?" (ibid., p. 57). If we subscribe to behaviouristic accounts of intelligence then anything can be said to be intelligent, including computers, animals, and rocks, without once addressing the issue of mental activity.[3]

Searle provides a thought experiment involving silicon brains which is revealing in light of this debate (1992, Chapter 3). Imagine that in the near future you have been diagnosed with an incurable brain disorder that will eventually lead to your death. As an experimental procedure, the doctors of the time suggest inserting silicon chips in your brain in order to forestall your eventual death. From this, there are three potential outcomes. The first is that to everyone's surprise, there is absolutely no change at all in your behaviour or your mental life – the causal powers of silicon match exactly those of the brain and you are able to continue your life exactly as you were before, except with a head full of silicon instead of grey matter. The second outcome is slightly different, there is no behavioural change in your actions, but your mental life slowly dissolves into nothingness. The doctors ask you questions and you can hear yourself answering but "you" are not in control of your body, it is functioning on autopilot as it were, and slowly, as more and more chips are added to your head, you disappear entirely, although no one can tell the difference. Third, you find that your mental life is exactly the same, you are

---

3 I will revisit the Turing Test and behaviourism more thoroughly in Chapter 2.

thinking and feeling and experiencing just as you were before, but you are not able to affect your body. It lies remotely on the bed with doctors surrounding you, lamenting the failed experiment as they remove your body from life support, claiming that you are brain dead as evidenced by the lack of behavioural response. What Searle hopes to show from this experiment is that the capacity of the brain to cause consciousness is distinct from its ability to bring about behaviour (ibid., p. 69). Behaviour is not a sufficient test for consciousness as we could have identical acts but very different conscious states – just as the behaviour of a theoretical zombie might be identical with that of a thinking human being, there is nonetheless a difference in what is going on in their brains. This difference is due in part to the property of "unified qualitative subjectivity" (Searle 2000, p. 557), which constitutes a significant portion the first-person ontology of consciousness. Zombies, thermostats, and computers are not conscious (and hence not intelligent) because they lack this quality (among others), and no degree of behavioural similarity can make up for this disparity.

1.2 Qualities of Consciousness

To explore that which so differentiates us from computers, there are three properties of consciousness that are most in need of explanation.[4] First is qualia, the feelings of what it is like to experience something; second, subjectivity, the first-person account of the world we have; third, unity, the singular conscious experience we have of the world, even though our sensory perceptions of it are wildly disparate. Far from being distinct, these three aspects "are logically interrelated … [and] different aspects of the same feature," (ibid., p. 560). They are at the core of what it is to be conscious. Dennett, however, challenges these three properties as he is skeptical of their ontologically subjective nature and does not see any need to involve any such notion in a discussion of consciousness (1991a, 2003). Dennett's critique is multi-faceted and he investigates qualia, subjectivity and unified consciousness throughout the latter chapters in *Consciousness Explained* in hopes of revealing absurdities and contradictions in those

---

4 There are many other properties in addition to these three. Searle (1992, 2000, 2004) lists up to a dozen different properties of consciousness that are, in their own ways, as interesting and in need of explanation. I focus on these as they are common within both Dennett and Searle's writings and they seem to provoke the most debate.

concepts. Qualia are denied existence while subjectivity, along with the unified feeling of consciousness, is deemed illusory (Dennett 1991a, Chapters 12-14). Dennett's philosophy is highly relevant in regards to posthumanism as neither he nor Searle refutes the claim that it is possible to (eventually) build conscious machines – and in fact Dennett sees little difference between us and such constructs[5] – but their interpretations of what it would mean for machines to be conscious, and how we should go about evaluating such a claim, are fundamentally at odds. If the goal of posthumanism is to offer the opportunity to completely recreate all the experiences of conscious humans, then we must know what constitutes consciousness and it is about these very properties that Dennett and Searle so strongly disagree.

1.2.1 Subjectivity

To begin, I turn to subjectivity and corresponding debates about personal identity. This is perhaps best viewed in accordance with Dennett's discussion of Hume and his claim that there is no intrinsic property of identity that guarantees personal identity across one's life. Hume states:

> For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. … If any one upon serious and unprejudic'd reflexion, thinks he has a different notion of *himself*, I must confess I can reason no longer with him. All I can allow him is, that he may be in the right as well as I, and that we are essentially different in this particular. He may, perhaps, perceive something simple and continu'd, which he calls *himself*; tho' I am certain there is no such principle in me. (Hume, p. 252)

For Hume, the feeling of self is confused and under closer scrutiny it evaporates into discrete instances of sense data. The self is a useful social fiction, best used in discussions of ethics and responsibility, but by no means is it a real property. Additionally, Hume has little use for arguments in favour of the existence of a soul or any other such property that could shoulder the burden of identity. Even if we were to propose such a property, it

---

5 "The CADBLIND Mark I has – I will allow it – a rather simple, impoverished color space with few of the associations or built-in biases of a human being's personal color space, but aside from this vast difference in dispositional complexity, there is no important difference. I could even put it this way: There is no qualitative difference between the CADBLIND's performance of such a task [i.e., recognizing different shades of red] and our own." (Dennett 1991, p. 374)

would need to perform some very difficult metaphysical work in order to match scientific models of the world. Dennett's conception of the self is similar in many respects to that of Hume's and Dennett further argues that if we lack a property to which we can attach identity ("pearls of soul-stuff") then this feeling of selfhood that many of us claim to have must equally be a phantasm (Dennett 1991a, Chapter 13). Searle's conception of subjectivity is, unsurprisingly, different from Dennett's and Hume's. First, while Searle is not claiming that there is, or is even searching for, an intrinsic property of identity, he does not describe the feeling of "what it feels like to me to be me," (Searle 1992, p. 252n3) as being in any way illusory or fictitious. There really is such a feeling and it stems from my having a particular physiology and biology, not an indestructible or essential property to which this feeling of identity of one's self is attached. Additionally, for Searle, this feeling is not an after-the-fact judgment nor is it a confused idea; it has its own ontology.

Furthermore, this feeling may well be identical with the feeling of self that others have about their own selves – we just do not know. What we do know, however, is that "we must postulate a self as something in addition to the experiences in order that we can make sense of the character of our experiences," (Searle 2004, p. 298). That is, there must be someone we can speak of who is having an experience in order to talk about there being an experience at all. This 'someone' does not solve the problem of personal identity because it does not assume an essential property of "I-ness" for the argument to work, but what it minimally requires is the acknowledgment that to have an experience, there must be 'someone' who has it. Dennett, who is cautious of subscribing to any view that hints of mysticism or an anti-scientific viewpoint, is understandably skeptical of these claims of subjectivity. He makes a misstep, however, when he links his doubts about the feeling of identity to the dubious property of intrinsic identity.[6] As a result of this misstep, he dismisses the notion that there is an "I" who has experiences or is conscious because he

---

6 This can be inferred from his characterization of beliefs about consciousness "as theorists' fictions similar to centres of mass, the Equator, and parallelograms of forces," (Dennett 2003, p. 20). The moral Dennett is trying to put across here is that in order to stay maximally objective, we cannot presume what is believed to exist as actually existing, and instead we must investigate only beliefs in order to "avoid any commitment to spurious data." (ibid., p. 3). The same rule holds for identity – it's fictitious until proven otherwise; all we can safely treat as evidence are beliefs of identity, not feelings of it (Dennett 1991).

sees it as residing in the same boat as the aforementioned property of identity. His dismissal is somewhat mitigated as he recognizes, like Searle, the need to have a 'self' if we are to make sense of the world. His solution is to propose that which would not be an actual self, but "a *representation* of a self" (Dennett 1991a, p. 429) or what he also calls a "center of narrative gravity". This representation tells the story of what happens to a body and in doing so creates the fiction of the self, which further enables us to speak of beliefs, to function in our environment, and so on. But make no mistake, there is no deeper structure or property of identity, there is only the illusion of it.

Dennett's view of identity is both powerful and contentious, however it is not yet clear that it is permissible, particularly because of his allegiance to Hume and his claim that our perceptions are received in atomistic instances. "But we know that that is wrong.... The Gestalt psychologists gave us a lot evidence for this nonatomistic but rather holistic character of our perceptual experiences," (Searle 2004, p. 298). Even before our sense organs transmit data to our brains, there is a capacity to mark and organize that data as "mine". For Searle, it is the self that performs such an operation and he describes it as analogous, in many ways, to a point-of-view that allows visual data to be interpreted. The self, like a point-of-view, is prior to the sensory information received by the brain. More importantly, just as visual information must be seen from a perspective if it is to be understood, Searle's notion of the self is such that it precedes all acts of perception and consciousness and allows that information to be incorporated.[7] It should be clear that Searle and Dennett, while coming very close in their conception of subjectivity and the self as a means to further interaction with the world, in actuality could not be more different. Searle affirms the existence, or at least argues for, a capacity for identity, which is revealed by the feeling of it, that is prior to the belief of the same. Meanwhile Dennett treats the feeling of self, like the feeling of consciousness, as an illusion and considers only beliefs about identity and consciousness as relevant to proper study, noting that deeper claims about the existence of a self based on feelings are potentially spurious (Dennett 2003).

---

7 But the self is more than just this. "It has to be an entity, such that one and the same entity has consciousness, perception, rationality, the capacity to engage in action, and the capacity to organize perceptions and reason, so as to perform voluntary actions on the presupposition of freedom. If you have got all of that, you have a self." (Searle 2004, p. 297).

Searle himself is uncertain that his solution to the problem of the self is sufficient (2004) but, as I read him, it seems that identity, or the ability to create such a construct, could be safely located in the Background, the non-conscious abilities of the brain that shape our conscious and intentional worlds. As such, his discussion of intentionality and its relationship to the Background can be adapted to answer some questions about the self and identity. Searle states:

> I can, for example, be committed to the proposition that objects are solid, without in any way, implicitly or explicitly, having any belief or conviction to that effect. Well then, what is the sense of commitment involved? At least this: I cannot, consistently with my behavior, deny that proposition. I cannot, while sitting in this chair, leaning on this desk, and resting my feet on this chair, consistently deny that objects are solid, because my behavior presupposes the solidity of these objects. It is in that sense that my intentional behavior, a manifestation of my Background capacities, commits me to the proposition that objects are solid, even though I need have formed no belief regarding the solidity of objects. (Searle 1992, p. 185)

I think that identity, or the feeling of self, is amenable to this interpretation. Without thinking of it, or having beliefs, or forming representations, about my identity, I nonetheless act as a person with a unique identity and this action furthers my progress in the world. This does not mean that I actually am such a being, but I cannot seriously begin challenging that notion in practice if I expect my interaction with the world to remain at all similar.[8] This poses a challenge to Dennett's heterophenomenology as he is interested in the beliefs people have regarding their conscious experiences and not the experiences as those are subjective and therefore in doubt. If so, then for Searle, the capacity for identity is more fundamental than belief and if so then the fictional account of identity given by Dennett is weakened. The *belief* of the self is secondary to the investigation of its relationship to consciousness as many non-pathological human beings may never scrutinize their feelings of selfhood and so not form any corresponding beliefs. Nonetheless, they continue to function as though they were unique individuals. We can,

---

8 It is interesting here to note that Sacks (1990) considers several cases in which his patients have either lost, or never had, the capacity for identity. He theorizes that people with "super-Tourette's" or severe memory damage exhibit behaviour that could be described as indicating a lack of self. Respectively, his patients with these conditions function as mimes and mimics of others' actions (ibid., p. 124) or they are incapable of linear or complex thought (ibid., p. 111-15). The belief of identity never arises in these cases because the capacity for identity has been damaged; as such I think this lends support for Searle's argument for the actual, and not fictional, existence of identity.

of course, draw our attention to our feeling of self, and then form a belief about it, but there must first be the feeling of identity, which rests upon the Background capacity for the same. Subsequently, Dennett's claim[9] that his heterophenomenological theory, which assumes the validity of objective scientific methodologies, can fully explain consciousness seems boastful. Searle is able to meet his call for evidence of the inadequacies of epistemic objective science; there is something additional to beliefs of identity – namely the need to have a constraint on sensory information in order for understanding to take place, which is expressed by the feeling of subjectivity – that Dennett's philosophy has not addressed (and, perhaps, cannot address). Accordingly, I view Dennett as premature in his disavowal of the subjective nature of consciousness, as Searle has revealed, or argued for the existence of, subjective features of identity that are outside the reach of Dennett's heterophenomenology.

By situating the problem within the categories delineated by Searle – epistemic and ontological objectivity and subjectivity – I think we can have it both ways. We can have a non-mystical conception of consciousness, complete with first-person subjectivity and the corresponding feeling of identity, without subscribing to a property of intrinsic identity. The idea of an essential self is problematic, as Hume and Dennett both realize, yet it does not follow that the feeling of subjectivity is equally so. There can be a capacity for a "me" who feels cold (and not just exhibits the behaviour of being cold) when it is chilly in my apartment, without requiring a corresponding property of "me-ness". A Background capacity for identity could help solve this puzzle by positing identity as an ontologically subjective feeling that emerges from the specific physical properties of the brain, but remains irreducible to third-person accounts. This is a less satisfying concept of subjectivity than one that would guarantee our respective uniqueness across time and space, but it is preferable to Dennett's as his view is hindered by the oversimplification and eventual denial of identity and subjectivity.

---

9 "I am urging that the prevailing methodology of scientific investigation on human consciousness is not only sound, but readily extendable in non-revolutionary ways to incorporate all the purported exotica and hard cases of human subjectivity. I want to put the burden of proof on those who insist that third-person science is incapable of grasping the nettle of consciousness." (Dennett 2003, p. 22-23).

1.2.2 Qualia

As has been seen, Dennett and Searle subscribe to very different philosophies and so it is no surprise that their conceptions of consciousness and qualia are correspondingly dissimilar. And so, it would be remiss of me not to mention that for Searle, "'qualia' is just a plural name for conscious states. Because 'consciousness' and 'qualia' are coextensive there seems no point introducing a special term," (Searle 2000, p. 561). Additionally, conscious states include acts of thinking as well as acts of perceiving and so Searle assigns no special role to qualia and its relationship to perception. There are qualitative feelings attached to thinking just as there are to seeing – they are both subjective experiences.[10] Although Dennett does not address specifically this notion of qualia – he views it more traditionally as a feeling associated with perception and not thought – his critique of qualia is nonetheless worthy of discussion as it raises some interesting problems regarding the supposed existence of qualitative states – a discussion that is not uncommon in artificial intelligence research. Furthermore, such a discussion will lead us to what is at the heart of the issue for Dennett and Searle (and for posthumanism) and it is not the hypothetical existence of some quality or property of consciousness (although Dennett (1991a) spends a great deal of time formulating the argument in this way), but the very nature and existence of consciousness and how we can be sure of our knowledge of it. Qualia, then, is perhaps best read as an epistemological issue rather than an ontological one – what do we know about our experiences and how can we justify that knowledge. As above, Searle's distinction between ontological and epistemic notions of objectivity will play a key role here in examining qualia and qualitative states.

To begin, I think it should be said that in several places Dennett's critique of some of the varieties of qualia is incredibly helpful in getting to the crux of the issue. He has little respect or time for epiphenomenal conceptions of qualia or dualist philosophies of mind and he is correct to draw attention to absurdities to which such theories give rise.[11] The success or failure of posthumanism cannot be evaluated if what is needed to be

---

10 See Searle 2000 and 2004.

11 Dennett 1991, Chapter 12.

duplicated is uncritically assumed to reside in the ethereal realm of the 'mind' or the 'soul' or is conceived as an impotent side effect of brain processes. In this light, Dennett's treatment of colour is most relevant here as it (and, as we shall see, all sensory experiences) is best conceived as a "lovely quality", a quality dependent upon an observer or group of observers. However, he emphasizes that it is incorrect to speak of these qualities as having any reality or existence without those observers (1991a, p. 379-380). Furthermore, there is no objective truth about colour or the visual systems that bring about colour experiences; a colour-blind person's experience of the world is as true as a person with full colour-vision. There are no (epistemically) objective measures of colour to which we can appeal as colour vision varies across and amongst species. The objective truth is that there are things in the world with reflective properties that, through the interplay of light and brains, are interpreted as having colour by various beings. If qualia are meant to be or rest upon a physical property, then we must cast them aside as there is little scientific evidence to support that view. So far, so good; Dennett supplies a great deal of detail in regards to colour vision and the variety of the systems that bring it about and suffice it to say that at this point, I do not disagree with his conception of colour or his disavowal of any notion of an intrinsic property of colour that *is* colour for all human beings across time and space. However, he is not content to leave it at this, and goes further to investigate the *experience* of colour, of which, he states

> You *seem* to be referring to a private shade of homogenous pink, but this is just how it seems to you, not how it is. That "quale" of yours is a character in good standing in the fictional world of your heterophenomenology, but what it turns out to be in the *real* world in your brain is just a complex of dispositions. (Dennett 1991a, p. 389, emphasis in original).

And so let us grant that Dennett is correct – all the qualia of sensory data are fictions – why then, does the belief in such feeling arise? The belief in colour arises as so: light strikes an object that has certain reflective properties, which alters the way the light is reflected into the eyes of a person. Such light causes a chemical change in the light-sensitive cones in the eye, which transmit that data to the brain, which reorganizes it in such a way as to cause a behavioural change in the person. The feeling of colour arises, presumably, in retrospect as a confused judgment: "I reacted that way because I like the colour red, which is some real thing". This is slightly oversimplified, and of course other

factors, such as memory, are involved as well. What is relevant is that for Dennett there is no quale of red involved in colour vision, nor would such a thing have any additional causal powers, it is an additional property, a fiction, used to describe or explain the behaviour after the fact. This holds for pain and other experiences as well. A sledgehammer comes down on my hand, it causes a variety of nerve endings to react, which cause my brain to release chemicals that pull my hand away and also stimulates my vocal cords to yell and moan. Retroactively I describe that behaviour as "Me being in pain" and say that "It seemed like I was in pain" as short-hand to explain my behaviour. We could even have a neuroscientist with some sort of portable MRI kit beaming magnetic waves at my brain and she could agree that, yes, the neural pathways in my brain were excited in such a way as to produce the behaviour which we commonly associate with "being in pain". There is no need to involve the use of subjective experiences such as "redness" or "pain" to explain my actions. We can arrive at a perfectly good theory of behaviour (and consciousness) without troubling ourselves over feelings or qualitative states.

One might respond by stating that "I *really* was in pain" and Dennett would nod and respond, "No, it *seemed* like you were in pain. You have made a few confused judgments about your behaviour".[12] The feeling of pain is like the "lovely qualities" mentioned above. If there were no human beings, there could be no loveliness to refer to. Such things are dependent upon a class of observers and are not found in the world in the way that mass and density are. Presumably, if you had a full understanding of the physics of the world, you could explain the behaviour associated with loveliness without once discussing feelings or experiences.[13] Accordingly, Dennett not only denies that there are intrinsic properties of politeness, but also the subjective experiences of pleasure, pain,

---

12 Dennett states: "You seem to think there's a difference between thinking (judging, deciding, being of the heartfelt opinion that) something seems pink to you and something *really seeming* pink to you. But there is no difference. There is no phenomenon as really seeming." (Dennett 1991, p. 364, emphasis in original).

13 Most likely by appealing to memes, which are gene-like things with an ability to shape our cultural, if not biological, evolution (Dennett 1991, p. 200-10). To be honest, I cannot figure out what memes are, as opposed to what they supposedly do and how they work. They are clearly supposed to do work in the brain, but the methods through which they shape thoughts and actions are not clearly spelled out.

etc., as they haven't any relevance to the "real" world. Note that while he denies qualia, he does admit that they do seem to exist. However, he states, "... it does *not* follow from this undeniable, universally attested fact that *there really is* phenomenology," (ibid., p. 366, emphasis in original). These qualitative states are fictions and they are not part of the world.[14] But why all this "seeming"? Why can't it actually *be*? Well, we can only *seem* to have qualitative experiences for Dennett because if we actually had such experiences then it would be tantamount to accepting that there are subjective, qualitative modalities of consciousness, which are, by his current definition, "not among the data of science, since they can never be properly verified by objective methods," (ibid., p. 70). You can *believe* that you have such experiences, and heterophenomenology is the tool Dennett proposes to investigate those beliefs, but there is nothing "underneath" those beliefs, no feelings of pain or of pleasure, just stimulus, behaviour and confused judgments.

I struggled with Dennett for a long time because he seems to be right. It does seem simpler, and more "scientific" to situate subjective feelings as confused judgments that do not require a special status within the world. However, a story eventually helped me overcome my confusion with Dennett and, in turn, side with Searle. One afternoon, while I was asleep, I had inadvertently left my window open and unbeknownst to me a wasp flew into my apartment. It must have landed on my bed because when I rolled over I was immediately awoken by a sharp, painful and burning sensation on my arm. In my sluggishly conscious state I reached over and tried to figure out what was causing the pain and the wasp stung me again, this time on my fingertips. Throughout the experience I was consciously struggling to form some sort of judgment about what was causing the pain ("Did I leave a pen on the bed?" I thought), but what is more, there was an undeniable feeling of pain that arose *before* I was fully awake. This *feeling* woke me and continued *while* I was trying to explain what was happening. I did not have time to form a "confused judgment" about the causes of my behaviour – my arm (and then fingertips) *hurt*. As such, I hold that there is an *immediacy* to pain, or colour, or any other experience, that is not adequately addressed by Dennett's philosophy. If all we did was

---

14 McDermott echoes Dennett here when he states, "A fictional world is a separate world even if the fictions are being produced by a machine. In important ways, fictions are exactly what qualia are, useful fictions with a grain of truth (because they are attached to real sensory events)." (McDermott, p. 157).

retroactively evaluate and explain our behaviour I could consider his theory to be more plausible or, alternatively, if all we did was evaluate *others'* behaviour Dennett's heterophenomenology could be considered a successful candidate for explaining consciousness, but it does not explain, to me, why I feel pain when I feel it.[15]

Additionally, Dennett's heterophenomenology is, in some ways, like an inversion of the Other Minds problem. Where the OMP assumes your own consciousness, it questions whether or not it can be known that *other* people are really conscious and not just incredibly sophisticated robots or zombies. Dennett holds that because it can *only* be objectively known that human beings *are* sophisticated robots, it follows that he is one as well. He disqualifies his own qualia on the grounds that others cannot know them. This is due to his reliance on a particular scientific methodology/world-view that can be codified as so:

(1) Everything in the world can be explained and known through an objective, scientific methodology (Dennett 2003, p. 23).
(2) If a phenomenon is not knowable and verifiable by scientific methodology, it must not exist (Dennett 1991a, p. 460-1).
(3) Science cannot objectively access subjective experiences. (ibid., p. 70).
(4) Science cannot explain subjective experiences, only the behaviour associated with such experiences (ibid., Chapter 12).
(5) Therefore, subjective feelings do not exist; there is only behaviour and confused judgments.

As in the preceding discussion of subjectivity, Searle's distinction between epistemic and ontological conceptions of objectivity and subjectivity are incredibly useful as they form direct challenges to (1) and (2). Searle states:

> People sometimes speak of the "scientific world-view" as if it were one view of how things are among others, as if there might be all sorts of world-views and "science" gave us one of them. In one way this is right; but in another way this is

15 In fact Dennett blocks the use of heterophenomenology from evaluating one's own experiences. It is, after all, the phenomenology of other people. It cannot be used on oneself because the reporter of the data is untrustworthy (2003).

> misleading and indeed suggests something false… [I]t suggests that science names a specific kind of ontology, as if there were a scientific reality that is different from, for example the reality of common sense. I think that is profoundly mistaken. …[S]cience does not name an ontological domain; it names rather a set of methods for finding out about anything at all that admits of systematic investigation. (2004, p. 302)

Dennett, of course, is aware of this – he is not ignorant of the methodologies of science – and his heterophenomenology is an attempt to bring epistemically objective methods into the analysis of beliefs, consciousness, etc. However, he, and to a lesser extent McDermott,[16] are working from the assumption that because qualitative states are impossible to prove (or disprove) given a scientific point-of-view, they must be declared fictional.[17] Yet we cannot deny a phenomenon because our point-of-view presupposes its non-existence and, moreover, the adoption of such a view does not make the data go away. Throughout history, humanity's epistemological tools have been insufficient in one manner or another yet the world itself was unaffected. Proposition (1) and (2) are far too bold as it can only be assumed that scientific methodologies can explain everything in the world – as a philosopher of science, Dennett should know this. The inaccessibility of qualia to external observers is troubling to scientific methodologies that require those observers, but not to qualia. As such, Dennett's reconfiguration of qualia as after the fact judgments is unnecessary and also objectionable (as I hope to have shown in my wasp example). My feeling of pain exists before and during the attempt to form judgments. His explanation does not capture my experience when I have it and his subsequent denial of that experience (because his methodologies cannot access it) is too strong by half. So, too, is his conclusion that there are no qualitative states from his sound dismissal of intrinsic properties of colour or pain. That there are no universal properties of pain does not mean that there are no experiences of pain, even if, to be fair, such experiences are not well understood. It behooves Dennett and McDermott to be more cautious in their

---

16 McDermott, while initially calling qualia and qualitative states a fiction, states this: "What she [Mary the colour scientist] learns by experiencing red is the 'ineffable' aspect of qualia, the part that is completely indescribable, seemingly arbitrary, and independent of information about neural activity." (p. 154). But she nonetheless learns something, just what is not exactly clear. I'm not claiming McDermott endorses qualia, but he does not rule it out in exactly the same way as Dennett does.

17 See Dennett 1991, p. 403-4.

discussion of qualitative experiences, not because there are subjective phenomena outside the reach of current scientific methodologies, but because there might be. Dennett has shown how strange and seemingly paradoxical qualia are; he has not shown that they do not exist.

Searle's discussion of qualitative states is limited because, as I noted above, he sees little point in distinguishing between qualia and consciousness. Still, if Searle's work is to function as a useful template by which we can judge posthuman projects then it must be able to provide some explanation as to how consciousness/qualia arise and how to proceed with our investigations. The above discussion on the irreducibility of consciousness is clearly relevant and I suggest that the feeling of pain, or of colour, arises in a similar manner. Like all conscious states, the feeling of pain is an emergent property that occurs through a variety of neuronal interactions that is further linked to more fundamental Background[18] capacities of identity, memory and so forth. Contrary to Dennett's theory, where the "feeling" of pain arises after the behaviour, for Searle it is through a combination of the stimulus and the causal powers of the brain that the feeling is created, and that it has its own powers to affect decisions, other feelings, and so on. It is, and remains, a product of a specific brain and with no mystical or epiphenomenal properties attached. However, the actual content of those experiences ("what it is like for *x*") is not reducible to those processes – its emergence is subjective. This is, perhaps, the most important difference between Searle and Dennett's work as Searle holds that qualia/consciousness are in the real world (not in a fictional one like McDermott and Dennett suggest). This is a very different claim than that of Dennett's as Searle is not saying that qualia can be known subjectively, but that they *are* ontologically subjective – they are constituents of consciousness that are immune to third-person characterizations. Their reality is as solid as that of a mountain or planet, but different in that the processes that create qualia do so in such a way as to mark them as impenetrably subjective. Even if there were an epistemically objective "cerebroscope" that could look at the organization of neurons in your brain it could not say "pain feels like *x* for you"; the qualitative

18 More will be said on the Background in Chapter 2.

experience of your pain *is* your pain and so inaccessible to others.[19] One could run countless simulations of virtual people exposed to varieties of stimuli who respond by saying "Pain is like *this*" and such responses nonetheless remain behavioural responses that do not capture the quale of pain.

Finally, such behavioural interpretations of pain, like heterophenomenological accounts, are akin to claiming that a pound of feathers is the same as a pound of lead. In one sense, yes, they are the same thing (a pound), but in another they are made up of different stuff and they act differently in the world (e.g., if dropped from a high tower one floats gently to the ground, the other plummets). In reality, they are very different objects that, through a standard of measurement, are grouped together and deemed identical. This objective measurement, however, tells us little about the properties under examination, as it begs the question of identity. It assumes these things are equal without proving that they are. This, of course, reveals a serious flaw in heterophenomenology. For Dennett the "primary data" in the quest to explain consciousness are utterances, which indicate beliefs of qualia, but not qualia *per se* (2003). This would allow the possibility for a computer that speaks the sentence "I believe I have subjective, qualitative experiences of colour" to have the same meaning and substance as a person who claims the same.[20] This is not to say that Dennett's heterophenomenology is entirely without merit; I think if we are to investigate consciousness we should evaluate people's beliefs about it, and so he is very much on the right path in that respect. But he does not go far enough as, by stopping at beliefs, he overlooks the existence of deeper data in order to meet the demands of his methodologies.[21] Such a move seems radically premature, particularly in consideration of the next section where I evaluate why qualitative states and experiences of consciousness are fundamentally irreducible to epistemic objective accounts.

---

19 More will be said in the next section on how this plays out, on what constitutes consciousness and how the experience of x is x in a fundamentally subjective way.

20 Claims of this sort will be evaluated in Chapter 2, particularly in respect to Searle's Chinese Room Argument. Note that I do not think this an unfair characterization of Dennett, insofar as he denies any feeling attached to colour or pain (1991, p. 374). I cannot see what the difference between the computer's statement and mine would be.

21 Of this Searle remarks "If we have a definition of science that forbids us from investigating part of the world, it is the definition that has to be changed and not the world." (1997, p. 114).

1.2.3 Unity

There is an experience of consciousness by most non-pathological human beings that is described as being continuous, linear and unified (i.e., a stream of consciousness). This account of consciousness differs radically from the biology of the brain and the body's senses, which come in fits and starts, and are often incomplete and scattered. This difference between the biology and the experience is an anomaly that has sent Dennett on his path towards the denial of consciousness and which keeps Searle on his track against the adequacy of scientific methodologies to fully capture the subjective components of consciousness. Similar in many respects to the above discussion of qualia, Dennett challenges the unity of consciousness on the grounds that the feeling of consciousness is an illusion.[22] He is convinced of this by neurological accounts of perception and memory that reveal the processes thought to contribute to our consciousness as extraordinarily dissimilar to our experience of them.[23] Here I must concede that Dennett is absolutely right, the processes that shape my conscious experiences of the world are not unified, continuous or in any sort of "stream". Nonetheless, while I can certainly focus my attention to a given object and accordant sense perception, I cannot isolate one experience without noting its relation to everything else going on in and around my body. I do not experience the world in a Humean barrage of distinct sensory ideas but in a holistic and unified manner. It is this experience that is in need of explanation. However, Dennett replies,

> To insist … that what is not *there* in the brain must nevertheless be *there* in the mind because it certainly *seems* to be there is pointless. …[I]t wouldn't be 'there' in any sense that could make a difference *to [your] own experiences*, let alone to [your] capacity to pass tests, press buttons, and so forth. (1991a, p. 362, emphasis in original).

There are two claims here, both worthy of analysis. The first, that because my experiences do not match my biology, the former must be fictional, and second, that even

---

22 Dennett states: "One of the most striking features of consciousness is its discontinuity… The discontinuity of consciousness is striking because of the apparent continuity of consciousness," (1991, p. 356).

23 Dennett 1991, Chapter 11.

if my experiences were not illusory, they would be ineffectual or epiphenomenal, are both bold claims and I shall deal with them in turn.

The first, of course, is similar in many respects to qualia, and so I will not return to that discussion. However, there is still the lingering possibility that one day scientists will be able to more accurately decompose consciousness into its constitutive parts and say, with finality, that consciousness is *x* and declare that subjectivity and qualia are, again, illusions and fictions. Searle offers one final defense against this line of attack and it is this that I turn to in defense of the unity of consciousness. He states:

> [Y]ou can't disprove the existence of conscious experiences by proving that they are only an appearance disguising the underlying reality, because *where consciousness is concerned the existence of the appearance is the reality.* If it seems to me exactly as if I am having conscious experiences, then I am having conscious experiences. … I might make various mistakes about my experiences, for example if I suffered from phantom limb pains. But whether reliably reported or not, the experience of feeling the pain is identical with the pain in a way that the experience of seeing a sunset is not identical with a sunset. (Searle 1997, p. 112, emphasis in original)

In other words, you can be wrong about your experiences (it is a trick of light, not a moving picture), but you cannot be wrong that you have one. This is not the same as claiming that we have privileged access or incorrigible knowledge about our inner self. And Searle points to several examples in which the knowledge of our own conscious lives is flawed while external parties have more authoritative explanations of what is actually transpiring[24] – this is, perhaps, why we have psychologists. There is a deeper claim here, however, which is immune to Dennett's critique: that our conscious experiences (including colour, pain and unity of experience) are constituted not just by sense data (and in some cases, not at all), but by our own subjectivity.[25] We cannot be wrong that we experience consciousness in a stream even if our brain physiology is scattered and uneven. To be conscious is to have a unified experience of consciousness, regardless of stimuli or external observers. This is, perhaps, why Searle (2000) has tied

---

24 See Searle 1992, p. 147-9.

25 "Where intentional mental states are concerned, the states themselves are constitutive of the seeming. The origin, in short, of our conviction of a special first-person authority lies simply in the fact that we cannot make the conventional reality/appearance distinction for appearances themselves." (Searle 1992, p. 146).

subjectivity, qualia, and unity together. In important ways they all presuppose and reinforce one another. It does not make sense to talk about consciousness without the experience of it and that experience is determined in many ways by the subjective, qualitative properties of our linear conception of consciousness. No third-person account of consciousness can remove or reduce this experience or make it anything but what it is experienced as being.

You can, as Dennett does, carve off bits of consciousness in hopes of removing the subjective elements, but in the end you will still have to deal with the experience of what it is like to have a conscious experience. Blind sight (Searle 1997), split brains (Gazzaniga), and masked priming (Dennett 2003) are examples of the insufficiency of this line of analysis. The conscious experience of what is happening differs from the sense data being received by the brain, even where the sense data result in behavioural changes that are inexplicable to the person exhibiting that behaviour. It is not, in these cases, that the people under investigation are *wrong* about their conscious experiences. Their experience is what needs to be explained and Searle, I think, is dead right in stating that third-person accounts cannot do that.

This leaves Dennett's second claim, which is that subjective states, if they did exist, would be impotent and entirely ineffectual in shaping my experience of the world or in my ability to engage it. Personally, of course, I think this automatically false; that I believe I experience consciousness one way and not another is one of the reasons why I prefer Searle's philosophy to Dennett's and is, in some small way, why this paper was written. If I am right in gauging my own motivations, my belief in subjective consciousness is not epiphenomenal, but Dennett can adequately deal with my beliefs – heterophenomenology is aimed for precisely that goal – and so saying I believe in consciousness as so construed above is little help. What is still at issue is why a continuum of consciousness might be useful to a species and it is that topic that I explore now. Consider frogs: these creatures, as most anyone can tell you, are jumpy and often seemingly nervous. They will often leap and swim away from shadows, regardless of whether those shadows are cast by hungry dogs, clouds or trees. People, on the other hand, most often do not jump at shadows – although on occasion we do, and historically we might have had good reason to do so – and we certainly do not jump when clouds

pass over the sun. What could explain this difference in behaviour? According to Dennett there is no qualitative difference (as there are no qualia) between the conscious experience of a frog and a person, or a machine labeling colours and a person performing the same task (Dennett 1992, p. 374), just different dispositions.

For instance, to the frog everything (presumably) is new and subsequently a surprise and a possible threat. Its senses perceive what is out there in the world in a very similar way as ours do, but there is no continuum of experience for the frog. We, on the other hand, do experience the world in a linear way, even if our sensory perception of the world is fragmentary and incomplete. This experience allows us to perceive that things happen in sequence, are predictable and ultimately understandable. The reason, then, why we do not jump at shadows is that we understand our environment and know that it is unlikely that the shadow being cast pertains to a predator, but instead to a harmless cloud many kilometres above. It is of course possible that a shadow could be indicative of a threat – if I was in the wilderness or in an unknown dark alley I might very well treat shadows as possible threats – and in such a case it is because my understanding and experience of the world has changed from what I remember and perceive as being normal. I recognize that my confidence and assuredness of safety is unwarranted because I cannot predict what will happen and so my behaviour adjusts accordingly. Frogs do not do this; they are in a permanent state of fright. Their experience of the world is radically different than ours not because our senses are radically different, but because our conscious experiences are.

Evolution only provides us with basic dispositions for or against something (for instance things that are larger than me are threatening, things that are smaller than me are not). The addition of a conscious experience of these things suddenly adds the potential for a more nuanced interaction with the world. For instance, in the case of the wasp, my conscious experience of the event allowed me to form explanations about the cause of my pain, in addition to reflecting on ways to stop it now and in the future. A conscious experience that is unified allows for all this to be done and enables subsequent analysis (and many other actions) as well. Contrary to what Dennett claims, this experience of the world as unified, linear continuum not only alters our experience of it (e.g., we are not afraid all the time) but it undeniably informs our ability to "pass tests, push buttons, and

so forth." (ibid., p. 362). Moreover, that there are buttons to push and tests to pass is resultant from our ability to recognize that we are safe and have time to think and build. If we did not think we were safe, and if we did not have some reason to justify that thought, we would be like the frog, constantly on edge. For frogs and spiders, Dennett's concept of consciousness is (perhaps) satisfactory. There is nothing more to "being scared" for those animals than exhibiting the behaviour of running away. For humans (and presumably many of our mammalian cousins) something more is needed. That Dennett is able to sit down and write his book is testament to the falsity of that book's main hypothesis – that the experience of consciousness is not what we think it is – if it were otherwise, I am not certain he could pick up his pen.

1.3 Conclusion

Dennett's philosophy of mind is often celebrated for unequivocally claiming that computers can be conscious and have similar, if not identical, experiences as humans have. On first glance, then, his project seems highly amenable to a posthuman project – the trouble of duplicating subjective experiences is no longer a problem because there *aren't* any subjective experiences to duplicate. But this does not satisfactorily explain why we do the things we do and why we are certain we have these experiences which, Dennett is quite right to point out, we cannot seem to explain without reference to our "inner" self. His subsequent removal of all subjectivity and references to qualitative states in his theory of consciousness is, however, premature and does not follow from the arguments he has put forward. More importantly for my purposes, his philosophy as so conceptualized is incapable of explaining some of the central components of what it is to be human. He can, therefore, offer little help for posthumanism.

What is interesting is that Dennett need not have taken this approach. He can have qualia, subjectivity and the full stable of conscious animals and still be a materialist, but by uncharitably linking qualitative states with intrinsic notions of qualia and identity, without adequately exploring alternative theories, he draws a caricature of consciousness and not a fair portrait. In doing so, he weakens his own theory by robbing it of explanatory force – why do we experience pains and colours *before* we are able to form judgments about them? Calling them all confused judgments might make for a simple

and elegant theory, but does not do justice to the complexity of our experiences. Finally, he may admit to no difference between "conscious" behaviour and consciousness itself, but he has not adequately performed the reduction – there are still more than a few leftovers in need of cleanup.

Searle's brand of consciousness is preferable to Dennett's precisely because we can have our cake and eat it, too. Contrary to Dennett's insinuations, Searle is not proposing an anti-scientific program – his frequent calls for a better science of the brain that exposes how consciousness arises support this, as does his conception of consciousness as an entirely physical, though emergent, property of the brain. Yet Searle does require the acknowledgment that once consciousness emerges, what it is *like* to have that particular conscious experience cannot be fully accessed by 3rd person methods; to do so would miss the point of having subjective experiences. This doesn't make consciousness inexplicable or insoluble, but instead respects the possibility of conscious states that are outside the realm of epistemic objectivity. Searle provides a program of study that preserves consciousness, allows for its examination and, as I shall explore in the next chapter, the possibility of its eventual duplication and manifestation in a new medium: the computer.

2.0 Computation

At the end of Chapter 1, I had reached a working definition of consciousness and completed a rudimentary exploration of a few of its facets. The aim of this chapter is to evaluate the suitability of this conception of consciousness to a posthuman project, which for my purposes is premised upon the assumption that the mind can be implemented in a digital medium. However, throughout *The Rediscovery of the Mind*, Searle challenges several central beliefs of artificial intelligence research including 1) Strong AI: the mind is a program, and 2) Computationalism: the brain works wholly, or largely, as a digital computer. This current chapter will focus on both of these criticisms as they entail what are arguably devastating consequences for any posthuman program. Searle's formulation of the mind as a product of a very specific type of biology with causal powers distinct from those of computers or programs poses a serious threat to posthumanism. Moreover, his characterization of computation as observer-relative and causally ineffective could very well preclude the possibility of a posthuman future ever being realized. McDermott, however, rejects this account of computation and presents what I think is a successful counter argument to Searle's definition of computation. It seems, then, that there is room for both Searle and McDermott to be correct: the former in that the mind is a product of biology and the latter in that computation is an objective process, which is implicated somehow in consciousness. To explore this rather unexpected possibility, a closer examination of Searle, Dennett and McDermott is in order.

2.1 Against Strong AI: The Chinese Room Argument

The first claim, Strong AI, is confronted in the guise of Searle's Chinese Room Argument (CRA), which unfolds as follows. Imagine that you find yourself in a room surrounded by many books containing clear instructions on how to manipulate strange symbols which you have never seen before or understand. A short while later, someone outside the room begins sliding slips of paper imprinted with those symbols underneath the door. You take these pieces of paper, look them up in your comprehensive library of symbol manipulating texts, and, following the books' instructions, write new symbols on a different piece of paper and slide it back under the door. What you have accomplished in performing such a strange and apparently meaningless process of syntax manipulation

is, to the person outside the room, to participate in a lengthy discussion on a variety of stimulating and difficult topics. However, this conversation was, entirely without your knowledge, partaken in Chinese, a language of which you are completely ignorant. The books that surrounded you were sets of rules from which you were able to derive pre-written responses and questions, which presented the compelling illusion that you, to the outside observer, were engaged in a perfectly pleasant and interesting dialogue. However, for you, you were only following rules, rules that were arbitrary and meaningless; you, as the syntax manipulator, were without even a tiny bit of understanding and, in fact, were clueless to the fact that a conversation had occurred.[26] From this it can be inferred that

> If you don't understand Chinese then no other computer could understand Chinese because no digital computer, just by virtue of running a program, has anything that you don't have. All that the computer has, as you have, is a formal program for manipulating uninterrupted Chinese symbols. (Searle 1984, p. 33).

In other words, you have, as Searle would say, "syntax but no semantics" (1980, para. 64). You, by virtue of the program you are implementing, are unable to attach meaning to the symbols being manipulated and the resultant outputs remain meaningless to you (if not to the people outside the room). This is because, as the CRA concludes, there is more to having a mind than just having the right sort of behaviour and computer programs, insofar as they are defined by rule-following behaviours, lack the ability to create intentional states. Such an ability requires causal powers beyond those of simple symbol manipulation. These powers, presumably, are found in the biology of the brain and differ from digital computers in that they are able to facilitate the creation of semantic content.[27] The doctrine of Strong AI folds under this criticism for if, as the supporters of Strong AI claim, our brains are computers and our minds are programs then you, as the person in the Chinese Room, should have had an understanding of the symbols equivalent to that of a speaker of Chinese. Yet you did not – you were unaware that a conversation

---

26 See Searle 1980 and 1984.

27 Of this, Searle states "It is not because I am the instantiation of a computer program that I am able to understand English and have other forms of intentionality .... but as far as we know it is because I am a certain sort of organism with a certain biological (i.e. chemical and physical) structure, and this structure, under certain conditions, is causally capable of producing perception, action, understanding, learning, and other intentional phenomena. ...[O]nly something with those causal powers could have that intentionality." (1980, para. 53)

was happening. From Searle's perspective, what is missing in Strong AI is an examination of intentionality, and without such an investigation, any posthuman project will necessarily be incomplete.

2.1.1 Intentionality

For Searle, intentional acts and states are defined by their having directedness or aboutness – they refer to some idea or object, and, in doing so, become *about* those objects or ideas. In this way we can speak of thoughts and actions as being meaningful for they refer to, or are about, something else. This is in distinction to conscious states, however, which have an awareness of other objects or ideas, but this awareness does not necessarily translate into meaning. According to Searle, one can have intentional states without being aware of them just as one can be conscious of something that, as we shall see, may or may not have any intentionality (Searle 1983, p. 1-3). To illustrate, consider that you can be conscious of a mood (e.g., depression) without that state being *about* anything – you are not unhappy about your job, love life, or anything else, but you are nonetheless depressed and you are aware of it. Alternatively, you can be depressed *because* of something (e.g., someone ran over your beloved pet dog). In both cases the awareness of the mood is the same – you know you are depressed and you feel the depression – but what differs is the intentional state. The depression in the first case is not about anything,[28] it does not refer to any additional object or thought, and so is not intentional (which is not to say that it would not affect other intentional states, because clearly it would). Meanwhile the second is intentional: it is about the death of your pet and your depression is directed to, or about, that fact. Note that an external viewer's interpretation of your behaviour is irrelevant as, in both cases, your actions are identical (you are weeping, sullen and clearly unhappy) but the meanings behind such emotions are very different. Naturally, someone could ascribe a deeper meaning to your depression in the first case, and they might think, from a distance, that you are suffering from heartbreak or some other malady, but that does not alter the fact that for you, your depression is utterly meaningless.

---

28 And let's assume for the sake of this example that there are no hidden, subconscious reasons for your depression.

This does not, however, mean that your depression is inexplicable – a psychiatrist or neuroscientist might surmise that your depression is due to a chemical imbalance in the brain and so would be able to explain your emotional state (and hopefully alter it), but it is important to stress that having a cause is not equivalent to having meaning. This is, I think, one of the key points behind Searle's CRA. Clearly the man in the locked room is causally interacting with the symbols as the resultant sentences owe their existence to his deft manipulations. To a person outside the room it seems as though the Chinese Room is an intelligent machine capable of understanding human language, engaging in witty répartée and so forth, thus supporting the belief that the room has a capacity for intentional acts. So, too, with many of our experiences of contemporary computers: they act intelligently, do what we tell them to do (for the most part), and so it seems as though they must be, in some way, intelligent. Yet, according to Searle, they are not. We are, instead, conflating the intentionality of the designers and programmers with the behaviour of our machines.

This insight leads directly to Searle's categorization of intentionality into three subsets: intrinsic, derived and as-if (1992, p. 78-80). The first type, which is not to be confused with the notion that objects in the world have inalienable, intrinsic semantics, refers to the physiological ability of some animals to have intentional mental states, complete with content. Human beings have intrinsic intentionality not because we have a special insight that tells us what things "really mean" but because our mental states are about objects and events in the world and so are meaningful, regardless of external interpretation; rocks, trees, and water molecules do not have such states, and neither, Searle supposes, do programs. Derived intentionality is best conceptualized as evidence of another's intrinsic intentionality. The hieroglyphics of ancient Egypt have derived intentionality because they reveal what other human beings thought, believed, and did. The glyphs themselves have no capacity for intentionality, they do not give the world meaning by existing, but they show that others who were capable of intentional acts used such symbols to do so and such symbols' intentionality is accordingly derivative. Finally, as-if intentionality – which really is not a form of intentionality at all – is ascribed to objects when it is convenient to treat them as intentional. We say plants believe the sun overhead to explain their movements in that direction, we say apples want to fall towards

the ground to explain gravity, and we say computers are angry at us when they crash or fail to work properly.

As-if intentionality is often used to smooth over the gaps in our understandings of how the world works. However, in none of the above cases do any of the actions reveal any intrinsic or derived intentionality. The actions all have causes but they can eventually be explained without the use of such words as "believe", "want" or "angry". This form of reduction cannot be accomplished with intrinsic intentionality for Searle's theory of intentionality follows a similar path as that of consciousness; it is an emergent phenomenon that is causally dependent upon a host of physical systems (of which, Searle readily admits, we understand very little), but is nonetheless irreducible to them. These physical processes are not, however, programs or symbols (for reasons which will be made clear further on).

In this light, a slight re-reading of the CRA would reveal not the total lack of intentionality in the Chinese Room for the man in the room clearly has, like most non-pathological human beings, the capacity for intentional acts, just as the books he reads have derivative intentionality. However, those forms of intentionality are not going to allow for a comprehension of Chinese (or anything else) by virtue of running through a few, or even very many, steps involving symbolic manipulations. As long as we remember this and understand why such words are, in these cases, metaphorical (they have either derived intentionality or we are speaking figuratively) then the facts about the Chinese Room should be clearer, but the hope for artificial intelligence remains in doubt. Dennett, however, refuses this consequence for AI research and his rebuttal takes two closely related forms: the Intentional Stance (1989) and the Systems Reply (1991a, 1996).

2.1.2 The Intentional Stance

The intentional stance, according to Dennett, is one of three predictive strategies (the other two are the physical and design stances) used to explain the future behaviour of systems and objects. In many cases we can predict future events by appealing to physics – astrophysicists know the sun will rise and set tomorrow because of the earth's revolution around the sun, which, barring some significant catastrophe, will continue on

for quite some time because of gravity, etc. – and in such cases we adopt the physical stance. Other situations, however, are more readily explained by referring to a designer – we know what a pop machine will do (dispense a can of pop in exchange for money) because it was designed to do such a thing; its behaviour is still causally tied to physics, of course, but there are more efficient means to predict its behaviour and this is done by adopting the design stance. Finally, there is the intentional stance and it can be used on systems such as animals, computers and people, provided that such systems are acting rationally (Dennett 1989, p. 23). Such terms as "want", "believe" and "desire", then, are not imbued with any explanatory value in and of themselves, but instead are used as short-hand to predict behaviour. For instance, it is not that someone believes (where "believes" refers to an intrinsic psychological/mental state with causal powers) Paul Martin is a good leader, and so votes for his party in an upcoming election, but instead because someone votes for Paul Martin, we suppose that they have such a belief. This attribution of a particular belief will allow us to make further predictions about this person's behaviour (perhaps the person is also a member of a provincial Liberal party and attends rallies of a moderate/progressive nature), but we need not posit some "deeper meaning" to their actions, or an internal, intrinsic mental state.

However, Dennett is not, like Churchland (1995), going so far here as to deny the existence of beliefs altogether. Instead, he reformulates beliefs as *abstracta*, analogous in many ways to the scientific concepts of centres of gravity or the equator (1991b). This does bring to mind further questions about the ontology of these abstracta – certainly they have great utility, they allow us to make predictions and facilitate scientific observation, but do they exist? Dennett suggests the answer can be found in the patterns we associate with such abstracta, in the case of beliefs, the behaviour of people and animals.[29] Those patterns are objectively real (e.g., people vote for different parties) and if we were to ignore that pattern, we would be missing something very real. So, in a more general sense, beliefs are real, but their existence is not to be attributed to mental qualities found "in the head".

---

29 See Dennett (1991b), p. 29.

With this move accomplished, Dennett has paved the way for the disbandment of any theory that would appeal to internal or intrinsic mental content as a means to explain behaviour – thus enabling a posthuman program to continue without need to explain such qualities. However, there are several important consequences that follow from this move, not the least of which is that the intentional stance evacuates any claim to first-person authority that we might have about our own thoughts. If we haven't intrinsic intentionality, just derived, we cannot appeal to our "inner selves" or some special property of intentionality to explain what we *really* meant – we do not have the authority, or power, to do such a thing. Second, Dennett argues that if we lack an interpretation about what an act or a behaviour signifies, then there simply is no way to determine its meaning and so the act *is* meaningless (1989, p. 300). Yet Dennett does not go so far as to claim that there is *no* meaning in the world; he envisions, instead, that human beings and all other intentional systems are rich with derived intentionality. Importantly, the only place from which this derivative intentionality can possibly come is via the interplay of genetic forces with environmental factors – in short, evolution.[30] Nonetheless, even this original intentionality (as Dennett calls it) is highly indeterminate; real meanings are dependent upon real functions but functions alter depending upon environment, species and observer. And so, put very crudely, we can either appeal to what other people thought our behaviour signified, or should observers be absent, we can theorize what Mother Nature had intended our species to do in a similar situation. In cases where there is no theory, or if there are many competing interpretations, then there simply is no meaning involved.

An immediate problem for such a theory is how does a process that has at best as-if intentionality become the foundation of all intentionality. Dennett suggests that "If we work out the rationales of these bits of organic genius, we will be left having to attribute – but not in any mysterious way – an emergent appreciation or recognition of those rationales to natural selection itself," (ibid., p. 317). But this raises another concern: if we have to perform such an attribution to natural selection, do we have to make a similar attribution to other blind and purposeless processes such as, for instance, the formation of

---

30 See Dennett (1989), Chapter 8

solar systems or of the universe? If not, why? Dennett admits that we are (or eventually will be) fully capable of describing evolutionary processes without mention of intentionality,[31] so why does he insist that we do so? His reply lies in his methodology of how the intentional stance should be used.

First, Dennett states that the intentional stance should only be used in cases where adopting the stance will add additional explanatory or predictive value. We do not say that a rock is an intentional system because claiming that a rock is solitary and immobile because it wants to be alone adds nothing to our predictive powers. We already knew what the rock was going to do: nothing. Adding beliefs and mental states to the rock does not increase our ability to theorize what it will do next. With intentional systems, however, we do not know what people will do next – the physical stance is far too complex for us to access in real time – so beliefs are attributed to people to explain their behaviour, and to predict what they would do next. For example, a person at a crosswalk will wait for the red light to change to a green "walk" sign because, I assume, they do not want to be hit by a car. After observing many people at many different times intersections there emerges a pattern of behaviour that is undeniably real and objective: green light means go; red light, stop. Ignoring this pattern and focusing only on the physical stance would not provide the same predictive abilities; it would seem as though people were just stopping and starting for no reason whatsoever – the coordinated activity of all these people would seem miraculous.

This, then, is the reason for Dennett placing the burden of original intentionality upon the unintelligent and blind processes of natural selection. Over time the patterns associated with evolution (tendency towards complexity, more sophisticated and varied modes of interaction with specific part of the world) are the only phenomena on which we can ground meaning in Dennett's very indeterminate and fluid sense. If we do not locate original intentionality with evolution, then we will be unable to prevent a vicious cycle of derived intentionality unless we are willing to ponder the existence of intrinsic and essential meanings. Locating original intentionality in this matrix of evolutionary and genetic forces allows Dennett, via his formulation of the Systems Reply, a means through

---

31 See Dennett (1989), p. 316-17.

which he can attack the Chinese Room and Searle's claim that syntax alone cannot constitute semantics.

2.1.3 The Systems Reply

The Systems Reply holds that while you do not understand Chinese, there is nonetheless understanding either in a subsystem of your brain, or in the Room itself.[32] A comparison is often invoked between neurons and brains: a single neuron certainly does not understand English, Chinese or anything else for that matter, but through a convergence of other neurons, electrochemicals and, perhaps, the right algorithm, brains eventually do. So, too, with the Chinese Room. *You*, as the symbol-shuffling automaton, do not understand Chinese but the entire room, through the ever increasing complexity of syntax and behaviour, does. Therefore, according to the Systems Reply, it is correct for Searle to assert that there is no understanding of Chinese in you, just as there is no understanding of English in your neurons, but incorrect to go so far as to judge the entire system as unintelligent. What the Systems Reply argues is that the Chinese Room has "*derived* intentionality, and that is the only kind of semantics there is," (Dennett 1989, p. 336) and what Searle is hoping to uncover is intrinsic intentionality, which Dennett hopes to have shown is implausible or unnecessary once intentionality has been grounded in evolution. If Dennett is right, then Searle's distinction between semantics and syntax loses its edge. Searle is arguing that there is no way for meaning to emerge from the system by virtue of behaviour and syntax while Dennett suggests that there is no deeper meaning in any system outside of interpretation and predictable patterns. The hardware of the system – be it biological or electronic – becomes irrelevant because the program's derived intentionality can be legitimately reduced to evolutionary processes, just as is our intentionality.

To illustrate, consider a simple reflex. For Searle such an action (e.g., involuntarily pulling your hand away from a hot surface) might be a Background ability, but it is not intentional. For Dennett, however, it will have intentionality derived from evolutionary forces that "selected" for organisms that had this reflex (presumably because there was survival value attached to such an act). A computer simulation of a reflex, if the

---

32 See Dennett 1991, or Searle 1984.

environmental conditions and other factors such as natural selection were duplicated, would have the same meaning. By extension, a program that exhibits the same behaviour as a human being would have the same meaning attached to it (and this is, of course, the foundation of the Turing Test) regardless of any appeal to first-person authority. Accordingly, the intentionality of the computer system remains undeniably derivative, but it is not derived from intrinsic human intentionality (as Searle holds), but from evolution, by way of human beings. By imbuing natural selection with a blind, purposeless (but original) intentionality, against which all other derivative forms of intentionality can be hinged and interpreted,[33] Dennett is able to void Searle's demand for an examination of intrinsic human intentionality and begin in earnest a research program aimed to realize artificial intelligence – and eventually posthumanism, as well.

There are, however, a few difficulties for Dennett that need to be addressed before I fully endorse his criticisms. First, Dennett's ambiguous conception of meaning seems to betray his own requirement for embeddedness in the world.[34] To explain, consider an instance of the Chinese Room where the conversation consists entirely about music, and so, unbeknownst to the person in the room, he is re-coding messages back and forth about how much he (and, according to Dennett, the system) loves music. The Systems Reply would hold that something in that machine has, it can be inferred by the complex behaviour exhibited, some understanding or knowledge about music. However, the person put into the Chinese Room was born entirely deaf and has utterly no conception of sound or auditory sensation at all. Neither, of course, does the Room. It has but 4 walls, a ceiling and floor, some pieces of paper and a human being biologically incapable of processing sound (and to ensure that no vibration at all is transmitted, assume the room exists within a complete vacuum). When the external inquisitors ask "What is your

---

33 "So if there is to be any original intentionality – original just in the sense of being derived from no other, ulterior source – the intentionality of natural selection deserves the honor. What is particularly satisfying about this is that we end the threatened regress of derivation with something of the right metaphysical sort: a blind and unrepresenting source of our sightful and and insightful powers of representation." (Dennett 1989, p. 318).

34 "The completion of the semantic interpretation of your beliefs, fixing the referents of your beliefs, requires, as in the case of the thermostat, facts about your actual embedding in the world. The principles, and problems, of interpretation that we discover when we attribute beliefs to people are the same principles and problems we discover when we look at the ludicrous, but blessedly simple, problem of attributing beliefs to a thermostat." (Dennett 1989, p. 32)

favourite song and why?" the man replies "'Helter Skelter' because it reminds me of my youth," and the inquisitors conclude that the man/Room must *really* be knowledgeable about music, and thus intelligent. However neither have ever heard such a song and the answer is, for both the Room and the man, entirely devoid of semantic content – there are only symbols, carefully planted by clever programmers foreseeing a question of this nature.

In response, Dennett could, if he was willing, attempt to extend the boundary of the System exponentially outwards to encompass not just the man in the Chinese Room, but all its programmers and so on, but this engenders its own problems. Such a move would make the Chinese Room's intentionality (and by extension all artificial intelligences) so wide and dispersed as to be wholly separate from its immediate environment. The machine itself could be blind and deaf and it would still be legitimate to attribute to it discriminations based on colour and sound. It appears as though meaning can be ascribed to a system without any requirement of experience or proximity with the issues at hand. Meaning is so derived and observer-dependent that, provided there are patterns to supply predictions and the presence of a theoretical framework to guide interpretation, anything goes. It is hard to locate the importance of physical embeddedness in this paradigm as it is easily dismissed. If so dismissed, then the claim that evolution is the shepherd of meaning becomes suspect. Evolutionary processes are constituted by the environment, which encompasses other species, the physical landscape and virtually everything else. If intelligent machines can so easily void this requirement, then what use can theories premised upon evolution be in guiding our interpretation of their behaviour? An evolutionary approach to understanding human behaviour no doubt has much use, but it does not seem that an extension to computational machines is equally warranted.

A theory of cognition that might give the notion of embeddedness its due is that of the Radical Embodied Cognition thesis, explored by Clark (2001), although by no means endorsed by him. Such a view holds that the environment and the body play integral roles in cognitive activities, such problem-solving. Furthermore, the body and environment are not just the media in which cognition takes place, but they "are intimately intermingled courtesy of processes of continuous reciprocal causation that criss-cross intuitive

boundaries," (Clark 2001, p. 128). Computational analyses of mental activity (necessarily?) lack this closely intertwined relationship as they too easily separate the body from the mind (the hardware from the software) and so are incapable of truly achieving mental phenomena.[35]

My second criticism of Dennett is related to the first in that his dismissal of first-person authority as a way to guarantee meaning is, I think, unwarranted in view of his evolutionary approach to intentionality. He seems to be confusing (or exploiting) the ambiguity between the unlikely possibility of something having an intrinsic meaning that persists throughout time in spite of varied interpretations, and the capacity of a system (such as a person) to have intentional states, rich with internal meaning. Searle, as a proponent of the latter view, argues that people can *really* mean something regardless of external validation.[36] My thought "I'm hungry", if it is neither vocalized nor acted upon (I'm not hungry enough to actually make myself something to eat, for instance), does mean something, even if there is no pattern of behaviour to observe that validates such an interpretation, or even if the opposite behaviour that one would expect *was* observed (no food was eaten). Surely in this case, some degree of first-person authority can be allowed (Or is it that I ascribe the belief "I am hungry" to myself the instant after I move to make something to eat?), but it is not clear that Dennett would grant this and if not this thought, then how many other internal monologues and thoughts become, if they are not part of observable, rational behaviour, utterly meaningless. This bleak picture of intentionality, like Dennett's austere theory of consciousness explored in Chapter 1, appears compelling at first glance, but under closer scrutiny it is clear that Dennett is not just asking us to give up claims of intrinsic intentionality and first-person authority, but rather he is asking us to do so with little justification. This leads to my third and final, criticism of Dennett's evolutionary theory of intentionality.

In his discussion of natural selection and evolution, Dennett stresses that evolution is non-intentional and non-conscious in the processes of "choosing"

---

35 This line of inquiry will be resumed in the next section, "Against Computationalism".

36 "My having [a given] belief is a matter of intrinsic intentionality, and not a matter of what anybody else chooses to say about me or how I behave or what sort of stance someone might adopt toward me." (Searle 1992, p. 155). There are important qualifications as Searle does not suggest that first-person authority is unlimited, and those are discussed below.

adaptations. However, he argues that it is nonetheless amenable to the intentional stance because of "the pattern that permits prediction and supports counterfactuals," (Dennett 1989, p. 316-17). For instance, because animals who lacked a given behaviour (e.g., avoidance of larger animals with sharp teeth) would have had limited survival potential compared to their more populous relatives with such a behaviour, we can assume that this behaviour aids in survival. Furthermore, we can theorize that animals with offspring who have this behaviour will be more successful in terms of reproduction and survival and feel secure in the opinion that the *purpose* and *function* of this behaviour (and the genes underneath) is to safeguard a species' survival. This lesson is extendable and becomes the means through which Dennett hopes to ground intentionality in evolution. However, these functional assignments, like the semantics we attribute to computer syntax, are dependent upon intentional observers.

> When we say the heart functions to pump blood, we are describing a causal sequence as a function, but such descriptions are always relative to our Intentionality, relative to our interests. So, we cannot eliminate Intentionality in general and replace it with teleological function, because teleological function only exists relative to intrinsic Intentionality. (Searle p. 183, 1991)

And so the reduction that Dennett requires for his criticisms against Searle to hold cannot be made. He cannot locate original intentionality with evolutionary processes when such processes are named by and dependent upon intentional beings. Instead Dennett must accept that there is absolutely no teleology in evolution, a species is never finished, it never has a purpose. The phrase "survival of the fittest" is neither a commandment nor a force, it is an observation that fuels a theory made by intentional beings. These observations are undeniably real and they are representative of causally efficacious physical systems, but they are meaningless, non-intentional and finally insufficient in the task of accounting for even Dennett's brand of indeterminate intentionality.

This is because the patterns associated with evolutionary adaptations are far too ambiguous, too susceptible to any interpretation, and ultimately remain meaningless unless one posits some form of real, directed intentionality underneath – a possibility he has explicitly denied by claiming that our genes are the original "Unmeant Meaners". Evolutionary processes are, like the forces that shape solar systems and galaxies, purposeless and blind. The patterns that unfold (e.g., increasing complexity,

diversification, niche abilities) are abstractions viewed from somewhere, are interpreted through and because of biological, social and ideological lenses. The question Dennett should be asking is how do organisms such as ourselves create meaning, how do we ascribe and designate the varied phenomena of the world as having significance and relevance. It is undeniable that we do such things, but it is unclear how these methods of creating content are accounted for in Dennett's theory, how the power to ascribe meaning develops if there is nothing to ascribe, no meaning, until after such an ascription is made or if the meaning exists beyond evolution's reach.

And so Dennett's attempt to locate original intentionality in non-thinking, non-representing evolutionary processes must fail. His theory allows for too many strange consequences and I can see no principled reason why if we are to view evolution as a purposeful process, why we should not envision the laws of physics as having a similar teleology – they too have patterns and predictable behaviour. Ultimately, all such physical processes have is as-if intentionality, which is, of course, none at all. This is not, however, to suggest that Dennett's exploration of evolution and its relationship to intentionality isn't worthwhile. Although it would undoubtedly cause him no small amount of displeasure, I think the thrust of his criticisms can be met by Searle's theory of the Background, which, it can be argued, implicitly incorporates elements of evolutionary selection.

Searle is cognisant of Dennett's criticisms, particularly the claim that first-person, intrinsic intentionality is unnecessary or unsound in light of an evolutionary explanation. Nonetheless, Searle endorses the existence of intrinsic intentionality, but with some important caveats. He states,

> All conscious intentionality – all thought, perceptions, understanding, etc. – determines conditions of satisfaction only relative to a set of capacities that are not and could not be part of that very conscious state. The actual content by itself is insufficient to determine the conditions of satisfaction. (Searle 1992, p. 189)

In order to meet these additional conditions of satisfaction, what is first required is a set of abilities and capacities that are non-intentional and non-representational – that are not, in short, mental. From these capacities, it is argued, mental states emerge and the infinite regress of intentionality is avoided by locating it in neurology of the brain. Importantly,

these neurological abilities are products of evolutionary changes which are arbitrary and undirected – they could have been different if our species' history was other than what it is. Consequently, our (including what ever number of other intentional species there are) brains/minds are capable of intrinsic intentionality which is inextricably linked to the biology of the brain, and is so constrained. The Background is the non-intentional foundation of our representations and interpretations of the natural world, which is itself a product of that same world. However, the Background need not be the way it is, different species have different mental capacities (or none at all) just as they have different physiological abilities. This challenges the notion of incorrigible access to our own mental thoughts – there is no guarantee of a perfect knowledge of our own inner states for why should there be – but does locate original, intrinsic intentionality in the emergent phenomena of the brain.[37] We can, then, appeal to our thoughts and the thoughts of others as having meaning, even though we may be incorrect about our own motivations or if others interpret our behaviour as having a different significance.

I see no reason, paradox or problem in suggesting that evolution could have facilitated the development of organisms that, through the complex interplay of genes and environment, were equipped with brains from which intrinsic/original intentionality emerges. Dennett speaks at length (1989) about the wonders of complexity that evolution is responsible for, but seems unwilling to allow it this one last move. It is entirely plausible to me (as, I think, it is to Searle) that blind evolutionary forces "selected" for organisms that were capable of mentally representing the world, that have subjective, qualitative states, that have some degree of first-person authority, even if this authority is by no means absolute. Dennett's consistent and vocal denial of this possibility reveals a bias on his end, a bias profoundly anti-biological in its allowance of meaning and understanding irrespective of experience or environment. He would have us deny the wonders of the human mind in order to achieve what would amount to a sophisticated machine capable of clever tricks. Dennett's philosophy, as so constituted, cannot be a suitable framework for posthumanism as it overlooks or undervalues key qualities of the

---

37 Why we have such a capacity is, of course, up for debate. It seems clear, however, that our intentional abilities provide certain advantages over species that lack it (e.g., plants). Whether non-intentional animals could have achieved our level of technological and social sophistication is a topic for another paper.

"human experience". Searle's philosophy, however, which is not anti-scientific, mystical or absurd, remains a rigorous alternative that respects scientific understandings of the world such as evolution and non-dualistic physical laws, but finds room for consciousness and intentionality. As such, Searle's Chinese Room Argument withstands Dennett's objections. What remains to be seen is if our intrinsic human intentionality, premised as it is upon a specific biology and evolutionary pathway, can be recreated in an entirely different medium. It is this question that has the most significance for posthumanism and artificial intelligence research as it calls into question the very foundations of these theories: computationalism.

2.2 Against Computationalism: Defining Computation

Having sufficiently dealt with Strong AI and the claim that the mind is akin to a program, Searle moves on to develop further critiques that, while sharing some similarities with the aforementioned argument, provide his strongest attacks against computationalism, which holds that all functions of the brain can be implemented by a digital computer. He summarizes these critiques, or difficulties, as follows: 1) "Syntax Is Not Intrinsic to Semantics", 2) "The Homunculus Fallacy", and 3) "Syntax Has No Causal Powers", and 4) "The Brain Does Not Do Information Processing" (Searle 1992, Chapter 9). While this section will not deal directly will all four of the above criticisms, they will significantly inform the following discussion. However, to begin I will compare and assess a variety of definitions of what constitutes a computer and also the process called computation. If posthumanism's success is dependent upon such machines, then we must know precisely what a computer is and how it works.

Built into many definitions of what is a digital computer is the axiom of multiple realizability (Searle 1992, p. 205-212). Multiple realizability is one of the key markers of a digital medium and is manifested in the ability of data to be transferred and (re) produced *ad infinitum*, for the medium itself, provided it is digital, bears no weight on what is being (re)produced. It is irrelevant, for instance, if a digital music file is created on one computer, then sent, transferred, or moved to a dozen different computers with a variety of hardware and software configurations – in the end the result is the same: an

identical copy.[38] This infinite and faithful reproducibility is a feature unique to digital media and is not shared with analogue environments. The desire, then, for cognitive scientists and posthumanist scholars to assert that the brain is a digital computer is clear. If the brain is a computer and the mind a program, then AI research could eventually lead to a program with human-level (or beyond) intelligence, complete with beliefs, intentionality, and consciousness – or the means to transfer the mind into a machine – without worrying about the biological intricacies of the brain. If the brain is not digital, then the entire premise of posthumanism is jeopardized; the move from the realm of the flesh to the digital would necessitate losses in quality and ability as faithful reproductions are an impossibility. Unless this discrepancy can somehow be resolved, we will be unable to know if, as our theoretical posthuman program was being realized, the digital transferring process was stripping away essential layers of humanity and sacrificing subjective experiences and understandings of the world for convincing behavioural simulations. The new posthumans' behaviour remains the same (from the Chinese Room and Weak AI arguments we know that any computer system can simulate any other physical system) but in actuality our future descendants are indistinct from video game characters: all the correct words and responses are there but they haven't any actual experiences – they are automatons. Therefore, we must know what a computer is, what is it capable of, and how that such a concept relates to the brain.

Searle, however, is skeptical there is a relationship at all between computers and the brain. Computationalist programs tell us very little about the intrinsic causal connections that bring about understanding, beliefs, or intentionality, because they depend upon outside observers to ascribe those qualities to it. This is because, as Searle states "[i]f computation is defined in terms of the assignment of syntax, then everything would be a digital computer, because any object whatever could have syntactical ascriptions made to it," (1992, p. 207-8). Computers are 'intelligent' because we interpret their behaviour as such, as opposed to human beings who are actually intelligent/ intentional even if there are no other people around. Similarly, because there is no objective physical process that can be appealed to in order to stabilize the notion of

---

38 Though the files differ temporally, it is the content that we are after and that remains indistinguishable from one copy to the next.

"digital computer" computer scientists must face the possibility that *anything* could be described as a computer, thus hindering the discovery of real properties of minds. Even if, as Searle suggests, the definition of computation is tightened up to prevent multiple realizability from spilling into infinite realizability by supplying a complete listing of computational causal powers, there is still the fact that, for Searle, computation is an observer-dependent phenomenon and not an objective feature of reality.

Of this, Searle states: "*...notions such as computation, algorithm, and program do not name intrinsic physical features of systems*. Computational states are not *discovered within* the physics, they are *assigned to* the physics" (1992, p. 210, emphasis in original). There can be *no* physical description of a computational state because such a state always requires an external observer (who is intentional) to ascribe and view those states as being computational. Without an observer such states are merely mechanical processes that carry no meaning, like the grinding of gears. Consequently, without an observer-independent characterization of what constitutes computation or the causal powers necessary to bring about cognition, the project of creating artificial intelligence will, and must, fail.

2.2.1 McDermott's Reply

Searle's definition of computation is, however, ardently contested by others in AI research and McDermott (2001) is one of its sharpest critics. McDermott attempts a redefinition that will block Searle's claim of computation's inherent observer-relativity and will cement it as a legitimately objective science, thus paving the way for a computational account of consciousness. To begin, he defines a computer as "a physical system whose outputs are a function of its inputs," (p. 169). This notion of function, which is the process through which the input is converted into the output, is, he readily admits, relative. For example, if the input of a system is 1, the output is 5, we can assume the function is ($f(x) = 5x$); if the input and output were different, then the function would be different as well. This is not, McDermott stresses, akin to claiming that computation is observer-relative or subjective, because the function, like the input and output, is independent of human observation. Whether or not someone knows what the function is, it exists regardless. In spite of this redefinition, the number of systems to which

computational descriptions are possible still seem quite large and McDermott must do more to supply us with limits if he is to meet Searle's charge of infinite realizability. After all, if it can be claimed (and McDermott does) that moving billiard balls compute their resting locations, then so too could it be asserted that molecules, atoms and perhaps even subatomic particles compute their trajectories. Indeed, he suggests this himself, stating that his definition of computation "is so general that every physical system can be construed as a computer, or several computers, or even an infinite set of computers," (p. 173-74). But this seems to move McDermott closer to the acceptance of Searle's critique than it does to discharge it. It is little help to meet the charge of observer-dependency by pronouncing that computation is universal in its application and so powerful as to be present in the workings of galaxies as well as atoms, without supplying a little more detail and just claiming that it is, irrespective of observation, objective.

Importantly, McDermott is aware of this objection and posits two additional rules that will, he argues, prevent computation from expanding indefinitely and implausibly. The first is continuity: a system can only be described as involving computation if "a small perturbation from the state makes a small difference in the output," (p. 173). In other words, the output of a computer must be connected somehow to its inputs – a small change in input should result in a small change in output, larger input changes should reveal correspondingly (and predictably) larger outputs. A wall or a solar system does not implement a program (although Searle (1992) suggests the former can be described as doing so) because a description of the wall's state has no bearing on its outputs or inputs. A small change in the wall's (apparent) program should alter what it computes, as evidenced by a change in both input and output, but Searle suggests no method to evaluate that computation. Moreover, it is difficult to imagine a means by which we could perform such an evaluation and so Searle's charge of extreme observer-relativity is somewhat constrained. Continuity alone, however, is still insufficient to cement computation as an authentic science for there is no guarantee of causality – the continuity between output and input could result from forces other than computation.

This leads McDermott to the second additional rule for computation: causality. A computational system should only be legitimately viewed as such if its behaviour is due to a computation, and not just amenable to such a description. McDermott supplies the

example of a furnace that is altered by a thermostat's computation of a specific function as evidence of causality. The thermostat, he argues, computes a function when it bends due to a specific ambient temperature in a particular environment, thus completing a circuit and allowing the furnace to raise the temperature. The relationship between the rise in temperature and the bimetallic strip completing the circuit is causally dependent on the strip's computation, and so can be described properly as a computer. To clarify, an illustration of computational causality might be helpful.

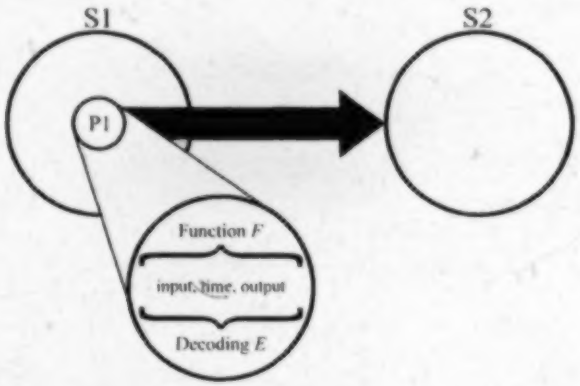

Figure 2.1. Computational causality

In the above diagram, the explanation of causality is as follows: there is a part (P1) of a system (S1) such that, under the decoding *E*, it computes the function *F* and is the only part responsible for influencing the other system in question (S2). In the case of the thermostat, the entire thermostat constitutes the first system (S1), and it is the bimetallic strip (P1) inside that bends (the output) according to the temperature of the room (the input) which computes the function *F* ("if x > y then true, else false" – if the temperature is greater than y, bend, if not, stay the same). Only the strip is responsible for the completion of the circuit which forces the furnace (S2) to raise the temperature, but it is useful as shorthand to attribute the computation of the function to the entire system.[39] This reveals an additional benefit of computation: compartmentalization; the system itself is not needed for functions to be computed, computation occurs independently of the

39 See McDermott, p. 177 for more examples of this sort.

whole.[40] This idea is closely related to the underpinnings of posthumanism, which work on the assumption that consciousness and intentionality, once they have been shown to be computational processes, can be separated somehow from the biological human system and reproduced faithfully elsewhere.

The strength of McDermott's reformulation of computation is that it effectively responds to Searle's criticisms, particularly those of observer-relativity and rule-following versus rule-governed behaviour.[41] While McDermott warns us against "philosophical decadence", such a revision nevertheless raises some challenging and important questions regarding computation. For instance, if functional descriptions are not observer-relative ascriptions, then what are they? Are algorithms and functions a heretofore undiscovered subset of physical laws? McDermott does not supply an answer to these questions as he himself admits that the study of computation is in its infancy as a science,[42] and much more research needs to be done to unpack the issues at hand. Nonetheless, McDermott does speak of a "computational realm" (p. 167) and while I mistakenly interpreted this at first as being a metaphor, now I am not so sure. McDermott seems to be situating computation as partaking in the fundamental fabric of reality, as residing in the same place where the law of gravity and other fundamental laws "live".

This move seems desirable and likely for McDermott as it would provide another means to guarantee the observer-independence of computation. Indeed, such an interpretation is strengthened by several examples supplied by McDermott. These examples are 1) "The VOR [vestibular-ocular reflex] evolved because the information it provides to the visual system is valuable," and 2) there is a line of cars with "a computer-controlled fuel-injection system, whose key element is a microchip known as the PowerPatsium, running the following program..." (McDermott, p. 180). Such examples cannot be reduced to purely physical accounts because such descriptions do not provide us with the same generalizations and predictive powers that computational explanations

---

40 See McDermott, p. 178.

41 See Searle's discussion (1992, Chapter 9) of the Homunculus Fallacy.

42 See McDermott, p. 216. I also think an exploration of what McDermott considers to be a legitimate science is in order, but that is outside the scope of this project.

enable. In the first case, the information the VOR provides explains why the physiology of the eye and brain evolved the way they did – a physical account cannot do this. In the second, the program which is causing the engine to start exists independently of the physical structure of the engine or the microchip – indeed, the program will continue to start engines in new cars with entirely different physical structures. Finally, McDermott has advanced a conception of computation that, while not entirely unproblematic, responds to many of the criticisms of that field and argues compellingly for its introduction into the scientific canon.

However, accepting computationalism as a legitimate science is not akin to embracing the idea that consciousness and intentionality are byproducts of computational processes (even if other features of the brain are). If anything, Searle provides us with a very clear warning about the ease with which computational assignments can be made to systems that are not best described in those terms (like walls and oceans, and including, perhaps, consciousness and intentionality). What McDermott provides us with are general guidelines by which we can investigate the suitability of those ascriptions without dooming computation as a human-centric fantasy. Moreover, even if we were to make the rather premature move to accept consciousness as resulting from computational processes, this does not immediately translate into success for a posthuman project or artificial intelligence research. Because, McDermott states, "there are parts of the body that probably do transmit information by changing their shape and hence the body's dynamics. ... we can't declare 'non-computational' effects irrelevant, because no state change is intrinsically noncomputational." (p. 179).

The importance of such a statement is that it raises furthers questions that cast doubt on the universality of functional accounts of consciousness. For instance, if mass, shape and density cannot be dismissed as non-computational, then can a similar claim be made about the unique biology/physiology of the human brain? Could the biological make-up of human beings allow our brains to transmit signals that give rise to conscious and intentional states because, as biological systems, they are susceptible to certain functional laws that non-biological systems are not? McDermott cannot be expected to answer these questions, but they do provide an opportunity to suggest that Searle and McDermott might both be correct. The former in that consciousness is a uniquely

biological phenomenon: the latter, by claiming that it is nonetheless due to computational processes that are observer-independent. Indeed, Dennett implies something much like this himself when he speaks of future artificial intelligence research producing powerful computational models of mind, adding: "*Perhaps one such model is psychologically real,*" (p. 86, emphasis mine).[43] From this passage, which sets McDermott apart from AI theorists of a more behaviourist/functionalist variety, there is a relatively short step to the idea that the psychologically real model will be a biological computation. Nonetheless, much more research needs to be done to explore computation *and* consciousness before generalizations between the two can be made, but McDermott and Searle theories combined could be a fruitful path to explore.

2.2.2 Diagnosis and Conclusion

There are, outside of Dennett and McDermott, a great number of alternative definitions and conceptions of computation. For a variety of reasons, I think they will be found lacking in respect to the development of a posthuman program, not because computation is observer-dependent, but because the whole endeavour so far has rested on the presumption that similar behaviour equals similar causes. This belief, expressed by McDermott when he claims that "mental terms [such as pain] can be defined so that they can be applied to systems without making any assumptions about what those systems are made of," (p. 25) is far too dismissive of the biological and physical foundations of those "mental terms". Nor is he the only proponent of this conception of functionalism. Moravec champions what he calls the "pattern-identity" position and the supposition that "the preservation of pattern and loss of substance is a normal part of every day life," (1988, p. 117) just as Dennett's intentional stance is, as we have seen, an exercise in pattern obsession. In one aspect they are all correct; there are patterns that emerge from biological processes, such as cell replication and so on, and the maintenance of these patterns are essential to the success and health of an organism. However, insofar as their focus is entirely on the pattern and not on the material, their projects will meet with failure as they ignore important components of our biology.

---

43 I am indebted to W. Cooper for alerting me to this passage.

Searle diagnoses this failure as a result not just of a belief in functionalism, but in a far deeper commitment to a methodology that adheres to a very curious chain of explanation. This chain involves three levels of explanation: the hardware level, which is analogous to the brain, followed by the software level, equivalent to the mind, and finally the knowledge/intentionality level, which is what AI is attempting to duplicate (Searle 1992, p. 215). As functionalism has it, the hardware level is irrelevant, we do not need to know how the hardware/brain works in order to understand how the intelligence arises. What AI researchers must do is create programs with patterns that match those found in the brain and they will have created programs with identical cognitive capacities as the mind, including consciousness and intentional states. Unfortunately, as Searle has argued, the program level is utterly impotent as it is observer-dependent and, as such, cannot offer a *physical/scientific* explanation of causality. This is because "[t]he implemented program has no causal powers other than those of the implementing medium because the program has no real existence, no ontology, beyond that of the implementing medium," (Searle 1992, p. 215). While McDermott has done much to curtail the strength of Searle's criticism regarding the observer-relativity of computation (and thus refuting the claim that it is causally impotent), he himself has recognized the importance of the "implementing medium".[44] In the case of intentionality and consciousness, it is the body which is both the source and the cause of the emergent patterns and the need, once it has been granted that patterns and programs are dependent in some ways upon the medium of the body, to champion pattern similarities found in minds and those in computers is lessened. Even if, to be generous, there are patterns in the brain that are indistinguishable from patterns in a computer, the latter does not explain the existence of the former, and this is precisely what we need to know if we are to evaluate the authenticity of a posthuman future. We must be certain that a computer can not just simulate the brain's patterns, but sufficiently duplicate their causes (and effects) so that our future posthumans will remain, in some way at least, human.

As such, there must be an acknowledgement that what has come first in nature, must also come first in attempts of duplication. We must begin with the materials we

---

44 See McDermott, p. 179.

know cause consciousness and intentionality, understand how these processes work, and then conduct experiments that will attempt to duplicate those processes in new, and perhaps longer-lasting, materials. If computationalists dismiss the importance of biology then they dismiss themselves from the possibility of explaining how and what consciousness is, and how we can create it for ourselves. We will not be able to realize a posthuman future by duplicating an abstract pattern or by programming a convincing simulation of intentional behaviour. We must know what and why the brain does what it does. Only then will we be able to duplicate, and not merely simulate, those same abilities.

In sum, the criticisms discussed above are sufficient to not only significantly curtail the dream of achieving artificial intelligence through programming alone, but also to challenge the entire science and rationale behind such beliefs. This has the additional consequence of reducing the possibility of a future posthuman existence. Or does it? If Searle is right then there will be very little success to be found in realizing a posthuman future if we continue along the paths set out for us by cognitive science. Yet this does not doom posthumanism to failure. Searle never once claims that machines cannot be intelligent – after all, we are (biological) machines and we are intelligent, so there is no logical impossibility in creating intelligent machines – he is opposed only to the view that intelligence can be achieved via computation alone.[45] Once functionalism is defeated, we can begin the project of building a better science of the brain in earnest. It is my hope that Searle, by providing a primarily biological account of consciousness, will supply a more solid foundation on which a posthuman project can be built.

---

45 "My own view is that only a machine could think, and indeed only very special kinds of machines, namely brains and machines that had the same causal powers as brains." (Searle 1980, para. 80)

3.0 Conclusion

In Chapter 1, I put forth several arguments that attempted to defend the view that consciousness is an emergent, subjective and, as far as we can tell, primarily biological phenomenon. So, too, in Chapter 2, did the view come across that intentionality is similarly emergent and organic. In addition to these insights, I argued against claims, such as Dennett's heterophenomenology and intentional stance, as well as McDermott's theory of computation, that consciousness and intentionality could be described entirely in third-person and computational accounts. Instead, I adopted and defended Searle's vision of these two phenomena being dependent on, but nonetheless irreducible to, the biology of the brain. Such a view requires the acknowledgment of inextricably subjective properties/modalities of consciousness and intentionality that are not accessible to an "objective" outside observer. This does not, however, result in a mystical or dualistic conception of the mind, but neither does it bode well for the success of a posthuman project, the likes of which has informed and shaped the purpose of this project. Typically, posthumanism is premised upon the abandonment of biology and its substitution with more durable materials, but if Searle is right, then the only means to ascertain the legitimacy of an artificial being's consciousness is to compare its foundations to the very specific causal powers of the brain that give rise to our intentionality and consciousness. Yet that is precisely of which we are most ignorant.

This leaves posthumanism at a very serious impasse. Either we can continue to build more impressive computer programs that appear to mimic (or duplicate) our own complex behaviour, and hope some version of functionalism ends up being true (but unverifiable). Or, we stall any such further research until we get a much better hold on the variations of consciousness and intentionality seen across a number of animal species and hope insight into the biology of our own brain tells us something about how such phenomena come about. Of course, neither option is mutually exclusive and both will most likely continue with great debate and (hopefully) success. What Searle's philosophy can do, I believe, is point out a way in which we might have the greatest chance of not just creating new forms intelligent beings, but preserve the entire range of human consciousness, as well.

This is in contradistinction to Moravec's vision of posthuman in "Mind Children"; he argues that if we are to survive into the distant future, if we are to guarantee that intelligence will not fade from existence with our species' very probable extinction, we must not tie ourselves to one conception of consciousness or being. We must be flexible and adaptable, abandoning the flesh and anything and everything else in order to safeguard the continued presence of intelligence in the universe.[46] This view of posthumanism is, however, the exact opposite of the one I would endorse.

I do not want facsimiles or otherwise poor copies of human beings populating a virtual world, nor do I think non-human, alien intelligences qualify properly as posthuman. Like Tipler's Omega Point, I think a posthuman future should include real human beings, or their digital equivalents. Survival in and of itself is not particularly compelling if the richness of humanity is lost in the struggle. It is not enough to just secure the continued presence of a cool alien intelligence in the universe; a posthuman project should allow the endurance of our human intelligence and consciousness. In order to do this, we must understand the roots and causes of our mental states by examining the biology and ontology of those states, not lauding the abstract behavioural similarities between computer simulations and human agents. Searle drives this point home again and again: we do not yet know what consciousness and intentionality are; we do not know what the processes and forces are that bring them about. It seems very likely that those forces are biological and that they based in the brain. As such, any project that would have us abandon our bodies and our brains without first understanding how they work, runs the risk of sacrificing unique and important phenomena such as emotion and consciousness for a dull, but lengthy, immortality. If we are to realize a posthuman future where the human, and not whatever is to follow, is to be identifiable and similar to biological human experiences, we must adopt a philosophy that respects that biology. Searle, as I have hoped to have shown in the preceding chapters, offers such an opportunity. While he supplies few easy answers, I think much success could be found by

---

46 "... we must die bit by bit if we are to succeed in the qualifying event – continued survival. In time, each of us will be a completely changed being, shaped more by external challenges than by our own desires. Our present memories and interests, having lost their relevance, will at best end up on a dusty archive, perhaps to be consulted once in a long while by a historian." (Moravec, p. 121)

adopting his theory of consciousness in the attempt to duplicate, in a digital medium, the entire gamut of human existence.

4.0 References

Bostrom, Nick. (2003). Are You Living in Computer Simulation? *Philosophical Quarterly, 53*(211), 243-255.

Churchland, Paul. (1995). *The Engine of Reason, the Seat of the Soul*. Cambridge: MIT Press.

Clark, Andy. (2001). *Mindware*. New York: Oxford University Press.

Clark, Andy. (2003). *Natural-Born Cyborgs*. Oxford: Oxford University Press.

Dennett, Daniel. (1989). *The Intentional Stance*. Cambridge, Massachusetts: The MIT Press.

Dennett, Daniel. (1991a). *Consciousness Explained*. Toronto: Little, Brown and Company.

Dennett, Daniel. (1991b). Real Patterns. *The Journal of Philosophy, 88*(1), 27-51.

Dennett, Daniel. (1996). *Kinds of Minds*. New York: BasicBooks.

Dennett, Daniel. (2003). Who's on First? Heterophenomenology Explained. *Journal of Consciousness Studies, 10*(9-10), 19-30.

Deutsch, David. (1997). *The Fabric of Reality*. New York: Penguin Books.

Flores, Fernando , & Winograd, Terry. (1986). *Understanding Computers and Cognition*. New York: Addison-Wesley.

Gazzaniga, Michael S. (1970). *The Bisected Brain*. New York: Appleton-Century-Crofts.

Hannan, Barbara. (1990). Critical Notice. *Mind, 99*(394), 291-297.

Hayles, N. Katherine. (1999). *How We Became Posthuman*. Chicago: The University of Chicago Press.

Heim, Michael. (1993). *The Metaphysics of Virtual Reality*. Oxford: Oxford University Press.

Herbert, Nick. (1993). *Elemental Mind: Human Consciousness and the New Physics*. New York: Penguin Group.

Hume, David. (1983). *An Enquiry Concerning the Principles of Morals*. Cambridge: Hackett Publishing Company.

Jacquette, Dale. (1988). Book Review of The Intentional Stance. *Mind, 97*(388), 619-624.

McDermott, Drew V. (2001). *Mind and Mechanism*. Cambridge: The MIT Press.

Sacks, Oliver. (1990). *The Man Who Mistook His Wife for a Hat*. New York: HarperCollins.

Schwartz, Jeffrey M., Stapp, Henry P., & Beauregard, Mario. (2004). The volitional influence of the mind on the brain, with special reference to emotional self-regulation. In M. Beauregard (Ed.), *Consciousness, Emotional Self-Regulation and the Brain*. Amsterdam: John Bejamins Publishing Company.

Searle, John. (1983). *Intentionality*. Cambridge: Cambridge University Press.

Searle, John. (1984). *Minds, Brains and Science*. Cambridge: Harvard University Press.

Searle, John. (1991). Perception and the Satisfactions of Intentionality. In E. Lepore & R. Van Gluick (Eds.), *John Searle and His Critics*. Cambridge: Basil Blackwell, Inc.

Searle, John. (1992). *The Rediscovery of the Mind*. Cambridge: The MIT Press.

Searle, John. (1995). *The Construction of Social Reality*. New York: The Free Press.

Searle, John. (1997). *The Mystery of Consciousness*. New York: The New York Review of Books.

Searle, John. (1998). *Mind, Language and Society*. New York: Basic Books.

Searle, John. (2000). Consciousness. *Annual Review of Neuroscience, 23*, 557-578.

Searle, John. (2002). Twenty-One Years in the Chinese Room. In J. Preston & M. Bishop (Eds.), *Views into the Chinese Room* (pp. 51-69). Oxford: Oxford University Press.

Searle, John. (2004). *Mind: A Brief Introduction*. Oxford: Oxford University Press.

Tipler, Frank J. (1994). *The Physics of Immortality*. New York: Anchor Books.

Turkle, Sherry. (1995). *Life on the Screen*. New York: Simon & Schuster.